28
1.9.2

JOURNAL OF CONSULTING PSYCHOLOGY

478243

Edited By
Laurance F. Shaffer
Teachers College, Columbia University

Associate Editors

Edgar A. Doll
Bellingham, Washington

Ann Magaret Garner
University of Illinois

William A. Hunt
Northwestern University

E. Lowell Kelly
University of Michigan

Morris Krugman
Board of Education
New York, N. Y.

Bertha M. Luckey
Board of Education
Cleveland, Ohio

Fred McKinney
University of Missouri

Anne Roe
New York, N. Y.

Carl R. Rogers
University of Chicago

R. Nevitt Sanford
University of California

Edward Joseph Shoben, Jr.
Teachers College
Columbia University

J. R. Wittenborn
Rutgers University

Managing Editor
Lorraine Bouthilet

BUREAU OF EDNL. & PSYL. [illegible]
LIBRARY

VOLUME 18

1954

Published Bimonthly By
The American Psychological Association, Inc.
1333 Sixteenth St., N. W.
Washington 6, D. C.

Printed At
Twenty-Three West Colorado Avenue
Colorado Springs, Colorado

Copyright, 1954, by the American Psychological Association, Inc.

Bureau Edn. Research
[illegible] COLLEGE
Dated 1.9.70
Acc. No. J268

Contents of Volume 18

Ainsworth, Patricia. See *Norman, R. D.*

Armstrong, Mary Ann Smith. Children's Responses to Animal and Human Figures in Thematic Pictures - - - - - - - - - - - - - - - 67

Auld, F., Jr. See *Murray, E. J.*

Baehr, G. O. The Comparative Effectiveness of Individual Psychotherapy, Group Psychotherapy, and a Combination of These Methods - - - - - - - - - - 179

Bandura, A. The Rorschach White Space Response and "Oppositional" Behavior - 17

Barron, F. A Correction - 150

Baughman, E. E. See *Belden, A. W.*

Bayard, Jean, and Pascal, G. R. Studies of Prognostic Criteria in the Case Records of Hospitalized Mental Patients: Affective Expression 122

Belden, A. W., and Baughman, E. E. The Effects of Figure-Ground Contrast upon Perception as Evaluated by a Modified Rorschach Technique - - - - - - 29

Bell, Dorothy M. See *Schofield, W.*

Bendig, A. W. Age, Sex, and the Manifest Anxiety Test - - - - - - - - - - - 16

Bendig, A. W., and Sprague, Janine. Rater Experience and the Reliability of Case History Ratings of Adjustment - - - - - - - - - - - - - - - - 207

Bendig, A. W. See *McCreary, Joyce B.*

Berger, S. See *Sacks, J. M.*

Bialick, I., and Hamlin, R. M. The Clinician as Judge: Details of Procedure in Judging Projective Material - - - - - - - - - - - - - - 239

Bills, R. E. Acceptance of Self as Measured by Interviews and the Index of Adjustment and Values - 22

Bills, R. E. Self Concepts and Rorschach Signs of Depression - - - - - - - - - 135

Bordin, E. S., Cutler, R. L., Dittmann, A. T., Harway, N. I., Rausch, H. L., and Rigler, D. Measurement Problems in Process Research on Psychotherapy - - - 79

Brackbill, G., and Little, K. B. MMPI Correlates of the Taylor Scale of Manifest Anxiety - - - - - - - - - - - - - - - - - 433

Branca, A. A. See *Eber, H. W.*

Brockway, Ann Lawler, Gleser, Goldine C., and Ulett, G. A. Rorschach Concepts of Normality - - - - - - - - - - - - - - - - - - - 259

Brown, Jean. See *Gelfand, L.*

Calden, G. See *Thurston, J. R.*

Calvin, A. D. See *Holtzman, W. H.*

Carr, A. C. Intra-Individual Consistency in Response to Tests of Varying Degrees of Ambiguity - 251

Charen, S. A Note on the Use of a Paper-and-Pencil Form of the MMPI *Hs* Scale for Hospital Use - - - - - - - - - - - - - - - - 344

Cochrane, C. M. See *Eber, H. W.*

Cohen, D. Rorschach Scores, Prognosis, and Course of Illness in Pulmonary Tuberculosis - - - - - - - - - - - - - - - - - - - 405

Cole, Mary E., Swensen, C. H., and Pascal, G. R. Prognostic Significance of Precipitating Stress in Mental Illness - - - - - - - - - - - 171

Cook, W. E. See *Woods, W. A.*

Corsini, R. J., and Uehling, H. F. A Cross Validation of Davidson's Rorschach Adjustment Scale - - - - - - - - - - - - - - - 277

Cowen, E. L. The "Negative Self Concept" as a Personality Measure - 138

Cromwell, R. L., and Lundy, R. M. Productivity of Clinical
Hypotheses on a Sentence Completion Test - - - - - - - - - - - - - - 421
Crowell, D. H., and Crowell, Doris C. Intelligence Test
Reliability for Cerebral Palsied Children - - - - - - - - - - - - - - 276
Crowell, Doris C. See *Crowell, D. H.*
Cummings, S. T. The Clinician as Judge: Judgments of Adjustment
from Rorschach Single-Card Performance - - - - - - - - - - - - 243
Cureton, E. E. Mental Age Equivalents for the Revised
Stanford-Binet Vocabulary Test - - - - - - - - - - - - - - - - 381
Cutler, R. L. See *Bordin, E. S.*
De Michele, J. H. The Interpretations of Anxiety by Various
Psychotherapeutic Schools - - - - - - - - - - - - - - - - - - 47
Dittmann, A. T. See *Bordin, E. S.*
Docter, R. F., and Winder, C. L. Delinquent vs. Nondelinquent
Performance on the Porteus Qualitative Maze Test - - - - - - - - 71
DuBois, P. H., and Hilden, A. H. A P Scale for the Rorschach:
A Methodological Study - - - - - - - - - - - - - - - - - - - 333
Eber, H. W., Cochrane, C. M., and Branca, A. A. Brief Intellectual Assessment of
Patients with Behavior Disorders - - - - - - - - - - - - - - - - 396
Edwards, B. See *Seeman, J.*
Eisen, Virginia W. See *Vane, Julia R.*
Feldman, Dorothy A., Pascal, G. R., and Swensen, C. H. Direction of Aggression
as a Prognostic Variable in Mental Illness - - - - - - - - - - - - 167
Fillmore, Arline R. See *Smith, L. M.*
Fleishman, M. The Discriminative Power of Szondi's Syndromes - - - - - - - - 89
Fonda, C. P. See *Schmidt, H. O.*
Gallagher, J. J. Test Indicators for Therapy Prognosis - - - - - - - - - - 409
Garfield, S. L., Heine, R. W., and Leventhal, M. An Evaluation of
Psychological Reports in a Clinical Setting - - - - - - - - - - - 281
Gelfand, L., Quarrington, B., Wideman, H., and Brown, Jean.
Inter-Judge Agreement on Traits Rated from the Rorschach - - - - - - - 471
Gibby, R. G., and Stotsky, B. A. Determinant Shift of
Psychoneurotic and Psychotic Patients - - - - - - - - - - - - - - 267
Gibby, R. G., Stotsky, B. A., Hiler, E. W., and Miller, D. R.
Validation of Rorschach Criteria for Predicting Duration of Therapy - - - - 185
Gibby, R. G., Stotsky, B. A., and Miller, D. R. Influence
of the Preceding Test on the Rorschach Protocol - - - - - - - - - - 463
Gleser, Goldine A. See *Brockway, Ann Lawler*
Goodman, M. See *McCue, Miriam*
Goodstein, L. D. Interrelationships among Several Measures
of Anxiety and Hostility - - - - - - - - - - - - - - - - - - 35
Goodstein, L. D. Regional Differences in MMPI Responses
among Male College Students - - - - - - - - - - - - - - - - 437
Gough, H. G. Some Common Misconceptions about Neuroticism - - - - - - - 287
Grauer, D. Homosexuality in Paranoid Schizophrenia as
Revealed by the Rorschach Test - - - - - - - - - - - - - - - 459
Guze, S. B. See *Matarazzo, J. D.*
Haines, Miriam Stearns. Test Performance of Preschool Children
With and Without Organic Brain Pathology - - - - - - - - - - - 371
Hales, W. M. See *Rosen, A.*

Hamlin, R. M. The Clinician as Judge: Implications of a Series of Studies - 233
Hamlin, R. M. See *Bialick, I.*
Hanlon, T. E., Hofstaetter, P. R., and O'Connor, J. P. Congruence of Self and Ideal Self in Relation to Personality Adjustment - - - - - - - - - - - 215
Harrington, R. W. Maladaptive Responses to Frustration Predicted from Rorschach Color Responses - - - - - - - - - - - - - - - - - - - 455
Harway, N. I. See *Bordin, E. S.*
Hastings, D. W. See *Schofield, W.*
Hathaway, S. R. See *Schofield, W.*
Heine, R. W. See *Garfield, S. L.*
Hemphill, J. K. See *Sechrest, L. B.*
Herman, J. See *Singer, J. L.*
Hilden, A. H. See *DuBois, P. H.*
Hiler, E. W. See *Gibby, R. G.*
Hofstaetter, P. R. See *Hanlon, T. E.*
Holsopple, J. Q. See *Lorr, M.*
Holtzman, W. H., Iscoe, I., and Calvin, A. D. Rorschach Color Responses and Manifest Anxiety in College Women - - - - - - - - - 317
Howard, A. R., and Shoemaker, D. J. An Evaluation of the Memory-for-Designs Test - 266
Ingram, Winifred. Prediction of Aggression from the Rorschach - - - - - - - - - 23
Iscoe, I. See *Holtzman, W. H.; Levine, B.*
Jacobs, A. Responses of Normals and Mental Hospital Patients to Freudian Sexual Symbols - 454
Jenkins, R. L. See *Lorr, M.*
Jourard, S. M. Moral Indignation: A Correlate of Denied Dislike of Parents' Traits - 59
Jourard, S. M., and Secord, P. F. Body Size and Body-Cathexis - - - - - - - - - 184
Karlin, L. See *Schwartz, M. M.*
Kendall, E. The Validity of Taylor's Manifest Anxiety Scale - - - - - - - - - 429
Kenny, D. T. Transcendence Indices, Extent of Personality Factors in Fantasy Responses, and the Ambiguity of TAT Cards - - - - - - - 345
Kostlan, A. A Method for the Empirical Study of Psychodiagnosis - - - - - - - - 83
Kurtz, Josephine C., and Riggs, Margaret M. An Attempt to Influence the Rorschach Test by Means of a Peripheral Set - - - - - - - - - 465
Lavos, G. Sex Differences on the Revised Beta Examination - - - - - - - - 375
Lawrence, J. F. See *Williams, H. L.*
Leventhal, M. See *Garfield, S. L.*
Levin, H. The Influence of Fullness of Interview on the Reliability, Discriminability, and Validity of Interview Judgments - - - - - - - - - - 303
Levine, B., and Iscoe, I. A Comparison of Raven's Progressive Matrices (1938) with a Short Form of the Wechsler-Bellevue - - - - - - - - 10
Levine, M. See *Meltzoff, J.*
Levitt, E. E. A Note on the Welsh MMPI Anxiety Index - - - - - - - - - - 112
Levy, L. H. Sexual Symbolism: A Validity Study - - - - - - - - - - - - - 43
Lindzey, G., and Newburg, A. S. Thematic Apperception Test: A Tentative Appraisal of Some "Signs" of Anxiety - - - - - - - - - - - 389
Little, K. B., and Shneidman, E. S. The Validity of MMPI Interpretations - 425
Little, K. B. See *Brackbill, G.*

Lonstein, M. A Validation of a Bender-Gestalt Scoring System - - - - - - - - - 377
Lorr, M., Jenkins, R. L., and Holsopple, J. Q. Factors Descriptive of Chronic Schizophrenics Selected for the Operation of Prefrontal Lobotomy - - - 293
Lotsoff, Antoinette B. See *Rotter, J. B.*
Lovinger, E. Note on Sanders and Pacht's Investigation of Size Constancy of Clinical Groups - - - - - - - - - - - - - - - - - 226
Lundy, R. M. See *Cromwell, R. L.*
McCreary, Joyce B., and Bendig, A. W. Comparison of Two Forms of the Manifest Anxiety Scale - 206
McCue, Miriam Crowley. Patient Cancellations Following a Sudden Cancellation by the Therapist - - - - - - - - - - - - - - - 176
McCue, Miriam, Goodman, M., and Rosenthal, M. Failure to Return for Treatment in Tested and Nontested Clinic Patients - - - - - - - 280
McGlothlin, W. H. A Psychometric Study of Gambling - - - - - - - - - - - 145
McIntyre, C. J. Sex, Age, and Iconicity as Factors in Projective Film Tests - - - 337
McMurray, J. G. Rigidity in Conceptual Thinking in Exogenous and Endogenous Mentally Retarded Children - - - - - - - - - - - - 366
Maher, B. A., and Martin, A. W. Mosaic Productions in Cerebro-arteriosclerosis - 40
Martin, A. W., and Wiechers, J. E. Raven's Colored Progressive Matrices and the Wechsler Intelligence Scale for Children - - - - - - - - - 143
Martin, A. W. See *Maher, B. A.*
Matarazzo, J. D., Ulett, G. A., Guze, S. B., and Saslow, G. The Relationship between Anxiety Level and Several Measures of Intelligence - - 201
Matarazzo, J. D. See *Mensh, I. N.*
Meltzoff, J., and Levine, M. The Relationship between Motor and Cognitive Inhibition 355
Mensh, I. N., and Matarazzo, J. D. Rorschach Card Rejection in Psychodiagnosis - - 271
Miller, D. R. See *Gibby, R. G.*
Mosel, J. N., and Roberts, June B. The Comparability of Measures of Profile Similarity: An Empirical Study - - - - - - - - - - - - - 61
Murray, E. J., Auld, F., Jr., and White, Alice M. A Psychotherapy Case Showing Progress but No Decrease in the Discomfort-Relief Quotient - - - 349
Navran, L. A Rationally Derived MMPI Scale to Measure Dependence - - - - - 192
Navran, L., and Stauffacher, J. C. Social Desirability as a Factor in Edwards Personality Preference Schedule Performance - - - - - - - - 442
Newburg, A. S. See *Lindzey, G.*
Newton, R. L. The Clinician as Judge: Total Rorschachs and Clinical Case Material - 248
Norman, R. D., and Ainsworth, Patricia. The Relationships among Projection, Empathy, Reality, and Adjustment, Operationally Defined - - - - - - - - 53
O'Connor, J. P. See *Hanlon, T. E.*
Omwake, Katharine T. The Relation between Acceptance of Self and Acceptance of Others Shown by Three Personality Inventories - - - - - 443
Parker, J. W. Tactual-Kinesthetic Perception as a Technique for Diagnosing Brain Damage - - - - - - - - - - - - - - - - - - - 415
Pascal, G. R. See *Bayard, Jean; Cole, Mary E.; Feldman, Dorothy A.; Swensen, C. H., Jr.*
Peterson, D. R. The Diagnosis of Subclinical Schizophrenia - - - - - - - - - 198
Quarrington, B. See *Gelfand, L.*
Rafferty, Janet E. See *Rotter, J. B.*
Rausch, H. L. See *Bordin, E. S.*
Riggs, Margaret, M. See *Kurtz, Josephine C.*
Rigler, D. See *Bordin, E. S.*

Roberts, June B. See *Mosel, J. N.*
Roberts, L. K. The Failure of Some Rorschach Indices to Predict the Outcome of Psychotherapy - 96
Rosen, A., Hales, W. M., and Simon, W. Classification of "Suicidal" Patients - - - 359
Rosen, A. Detection of Suicidal Patients: An Example of Some Limitations in the Prediction of Infrequent Events - - - - - - - - - 397
Rosenthal, M. See *McCue, Miriam*
Rosenwald, A. K. An Assessment of the Current Role of Psychodiagnostic Testing - 311
Rotter, J. B., Rafferty, Janet E., and Lotsof, Antoinette B. The Validity of the Rotter Incomplete Sentences Blank: High School Form - - 105
Rubin, H. Validity of a Critical-Item Scale for Schizophrenia on the MMPI - - - 219
Sacks, J. M., and Berger, S. Group Therapy Techniques with Hospitalized Chronic Schizophrenic Patients - - - - - - - - - - - - - 297
Saslow, G. See *Matarazzo, J. D.*
Schmidt, H. O., Fonda, C. P., and Wesley, Elizabeth L. A Note on Consistency of Rigidity as a Personality Variable - - - - - - - - - - - 450
Schofield, W., Hathaway, S. R., Hastings, D. W., and Bell, Dorothy M. Prognostic Factors in Schizophrenia - - - - - - - - - - - - - - - - - - 155
Schwartz, M. M., and Karlin, L. A New Technique for Studying the Meaning of Performance on the Rosenzweig Picture-Frustration Study - - - 131
Sechrest, L. B., and Hemphill, J. K. Motivational Variables in the Assuming of Combat Obligation - - - - - - - - - - - - - - - - 113
Secord, P. F. See *Jourard, S. M.*
Seeman, J., and Edwards, B. A Therapeutic Approach to Reading Difficulties - - - 451
Shneidman, E. S. See *Little, K. B.*
Shoemaker, D. J. See *Howard, A. R.*
Simon, W. See *Rosen, A.*
Singer, J. L. Projected Familial Attitudes as a Function of Socioeconomic Status and Psychopathology - - - - - - - - - - - - - - - - - 99
Singer, J. L., and Herman, J. Motor and Fantasy Correlates of Rorschach Human Movement Responses - - - - - - - - - - - - - - - 325
Singer, J. L., and Spohn, H. E. Some Behavioral Correlates of Rorschach's Experience-Type - - - - - - - - - - - - - - - - - - - 1
Smith, L. M., and Fillmore, Arline R. The Ammons FRPV Test and the WISC for Remedial Reading Cases - - - - - - - - - - - - - - - - - 332
Spohn, H. E. See *Singer, J. L.*
Sprague, Janine. See *Bendig, A. W.*
Stanley, J. C. Gellerman's Complex Crossover Design - - - - - - - - - - - - 380
Stauffacher, J. C. See *Navran, L.*
Stopol, M. S. Rorschach Performance in Relation to Two Types of Stress - - - - 11
Stotsky, B. A. See *Gibby, R. G.*
Swensen, C. H. Jr., and Pascal, G. R. Prognostic Significance of Type of Onset of Mental Illness - 127
Swensen, C. H., Jr., and Pascal, G. R. Duration of Illness as a Prognostic Indicator in Mental Illness - - - - - - - - - - - - - - - - 363
Swensen, C. H. See *Cole, Mary E.; Feldman, Dorothy A.*
Teevan, R. C. Personality Correlates of Undergraduate Field of Specialization - - - 212
Thurston, J. R., and Calden, G. Intelligence Factors in Irregular Discharge among Tuberculosis Patients - - - - - - - - - - - - - - - 404
Uehling, H. F. See *Corsini, R. J.*
Ulett, G. A. See *Brockway, Ann Lawler; Matarazzo, J. D.*

Vane, Julia R. Implications of the Performance of
Delinquent Girls on the Rosenzweig Picture-Frustration Study - - - - - - - - 414
Vane, Julia R., and Eisen, Virginia W. Wechsler-Bellevue
Performance of Delinquent and Nondelinquent Girls - - - - - - - - - - - 221
Wagenheim, Lillian. The Effect of Childhood Diseases on IQ Variability - - - - - 354
Wesley, Elizabeth L. See *Schmidt, H. O.*
White, Alice M. See *Murray, E. J.*
Wideman, H. See *Gelfand, L.*
Wiechers, J. E. See *Martin, A. W.*
Williams, H. L., and Lawrence, J. F. Comparison of the
Rorschach and MMPI by Means of Factor Analysis - - - - - - - - - - - 193
Winder, C. L. See *Docter, R. F.*
Woods, W. A., and Cook, W. E. Proficiency in Drawing and
Placement of Hands in Drawings of the Human Figure - - - - - - - - - - 119
Zelen, S. L. Acceptance and Acceptability: An Examination of Social Reciprocity - - 316
Zimmer, H. Self-Acceptance and Its Relation to Conflict - - - - - - - - - - - 447
New Books and Tests - - - - - - - - - - - - - - - - 74, 151, 229, 307, 384, 472
Index to Volume 18 - 477

Journal of Consulting Psychology
Vol. 18, No. 1, 1954

478243

Some Behavioral Correlates of Rorschach's Experience-Type

Jerome L. Singer
Franklin D. Roosevelt VA Hospital, Montrose, N. Y.

and Herbert E. Spohn
Teachers College, Columbia University

Within the framework of the Rorschach method the experience-type, the ratio of human movement (M) to weighted color responses (*Sum C*), although central to Rorschach's thinking [15], is in actual practice rather difficult to employ, as Beck [1] has pointed out. Part of the difficulty stems perhaps from Rorschach's somewhat arbitrary juxtaposition of two types of response which result from rather disparate perceptual processes. The situation is ameliorated somewhat by the fact that Rorschach has presented fairly specific indications of the behavioral significance of [illegible] modes of response as well as of the overt behavior occurring in persons with a clear preponderance of one component of the ratio. As part of a program of research designed to integrate key Rorschach determinants with general theories of perception and thinking, one focus of this study will be on the correlates of Rorschach's experience-type as given in the M:*Sum C* ratio. Because of the interesting theoretical implications of the human movement response, however, primary emphasis in this study will be upon the M-component of Rorschach's ratio. The color responses will be considered primarily from the point of view of how their presence in a record modifies the observed relationship between M and overt behavior.

Largely on the basis of intuition and empirical observation Rorschach [15] proposed that a triadic relationship exists between overt motor activity, perception of movement, and imaginative processes. He indicated that perception of humans in motion on the inkblots was correlated on the one hand with "greater creative ability," "more 'inner' life," and on the other hand, with "stable, measured motility" [15, p. 78]. He specifically asserted that "Kinesthetic engrams, therefore, act as inhibitors of physical activity; motor activity inhibits kinesthetic engrams" [15, p. 80]. Schachtel has summarized this view of the kinesthetic (M) response very effectively in his thoughtful paper [illegible]:

> . . . the kinesthetic percept is accompanied by arrested, checked motor innervations. Instead of being carried out, these innervations lead to empathy with movement observed in the environment, or to kinesthetic imagery as in kinesthetic percepts in Rorschach's test or as in dreams where motor activity is suspended by sleep [16, p. 87].

The challenge of this triadic relationship lies in its ready application to theories and observations concerning the development of thought which stem from diverse investigators in the fields of psychology and psychiatry. Thus, Freud in presenting his concept of the *reality-principle* as a key phase of ego development also pointed to the relationship between thought and restraint of motor discharge. He described thought as "endowed with qualities which made it possible for the mental apparatus to support increased tension during a delay in the process of discharge. It is essentially an experimental way of acting . . ." [4, p. 16]. Other psychoanalytic theorists [2, 3, 13] have elaborated on Freud's views but generally seem to concur in relating the origin of thought to a delay or inhibition of immediate need gratification through motor activity.

---

[1] Reviewed in the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions do not necessarily reflect the opinion or policy of the Veterans Administration. The authors are indebted to clinical psychologist trainees Jack Herman and Paul Pere as well as to the staff of the Psychology Service for assistance in various phases of the investigation.

Somewhat similar theories supported with some empirical data are presented by Piaget [11], Lewin [8], and Murphy [10]. All have in common a conception that the development of imagination and planful fantasy are concomitants of the increased restriction of the growing child's motility as well as of the enforced socialization of his egocentric speech patterns. To the extent that Rorschach's *M* represents fantasy activity, developmental findings with the inkblots are consistent with these theories. Thus, *M* first appears as a rule in children's records at age six, at a time when school attendance markedly curbs motility and enforces socialization [7, 19]. Differences between schizoid and normal children [19] or between disturbed children producing different amounts of *M* [7] also appear to be in accord with these theories.

With respect to the actual perceptual process in the production of Rorschach movement responses, experimental literature in perception and physiological psychology offers some clues [20, 21, 23]. Werner [21], for example, has hypothesized th[illegible]mon dynamic property, *tonicity*, underlies both sensory and motor responses. Available tonic energy may be released through body movement or may increase tonicity in a sensory area bringing about spatial displacement and the illusion of motion. Experimental support for this point of view with particular implications for Rorschach theory has come in a series of experimental studies [5, 23] and in comparisons of the perceptual and Rorschach behavior of feebleminded children who represent contrasting tendencies to motor activity [21, 22]. Thus, the hyperactive exogenous feebleminded showed fewer Rorschach movement responses than did the motorically phlegmatic endogenous feebleminded [22]. Evidence that the number of *M* produced on the inkblots by normals increases following a period of motor inhibition [9, 18] also supports Werner's hypothesis. Of particular interest for the present study was the finding that the number of *M* initially produced by subjects was positively correlated with their subsequent ability to inhibit motor response [9].

Concern in the present investigation will be with *M* in its hypothetical relation to motor behavior rather than with its presumptive correlates in fantasy and imagination. To ascertain the generality of findings obtained previously with normals, this study was extended to include a psychopathological group, i.e., schizophrenic males. In addition, a situation designed to elicit evidence of spontaneous motor activity and environmental responsiveness was utilized, since previous research [9, 18] had been limited to fairly specific motor inhibition tasks. It was felt that on theoretical grounds the presence [illegible] *M* responses in the Rorschach records of schizophrenics would bear the same general relation to motor inhibition that was observed in studies with normals. In the sphere of spontaneous motor activity, however, the color responses in the *M*: *Sum C* ratio might be expected to play a role in determining the degree of responsiveness. This study, moreover, involves the behavioral correlates of Rorschach's *M* and *C* categories as elicited in standard clinical administrations of the test rather than in the quantitative changes in these responses following inhibition as in the earlier work [9, 18]. For this reason the present study may be expected to throw some light more directly upon clinical interpretations of the response categories oriented toward description of personality structure.

The Experiment

Hypotheses. The following hypotheses were formulated:

1. Schizophrenic groups chosen as extremes of introversive or extratensive experience-types will differ significantly in two behavioral situations. Introversives ($M > C$) will show longer inhibition times in a motor inhibition task and less motor activity during a waiting period than extratensives ($C > M$).

2. Schizophrenic *S*s who have produced several human movement responses in their Rorschach protocols (High-*M* group) will show longer inhibition times and less waiting-room activity than an equated schizophrenic group who have produced relatively few Rorschach *M* (Low-*M* group).

3. The greater the number of Rorschach color responses in the patients' records, the greater will be the tendency for environmental responsiveness through motor activity during the waiting period. If, however, Hypothesis 2

4787 93

is confirmed, the quantity of color responses may be expected to exercise a modifying rather than a determining influence upon the amount of motor activity.

Subjects. The *S*s were 50 male veteran patients officially diagnosed as schizophrenic. They were divided into two groups of 25 *S*s each on the basis of the *M* responses in their Rorschach protocols.[2] Those producing more than one *M* were designated as the High-*M* group while those whose records yielded one *M* (usually Card III) or less composed the Low-*M* group. These groups were equated for other variables which might conceivably be related to the experimental situation, e.g., age, duration of hospitalization, IQ, and Rorschach scores. Table 1 presents comparisons of the means for all groups on certain pertinent variables. Others which showed low incidence are omitted as are percentages. For the High- and Low-*M* groups only the difference in *M* was significant. Both groups contained approximately equal numbers of the various subtypes of schizophrenia, predominantly paranoids and hebephrenics in this sample.[3]

In order to test the hypothesis concerning introversion and extratensiveness, two equated groups of *S*s chosen as the extremes of the *M*: *Sum C* ratio for the patient population were also employed. These groups were chosen to exclude *S*s with "coarctated" or "ambiequal" ratios. These groups consisted of 19 introversive *S*s and 16 extratensives who were equated on all variables except *M* and *C*. Data for these groups are also presented in Table 1. Only the differences in *M, CF, C,* and *Sum C* are significant.

Procedure. The *S*s underwent the following sequence of procedures:

1. Waiting-room observation

Table 1
Group Means for High- and Low-*M* and Introversive-Extratensive Groups on Age, IQ, and Pertinent Rorschach Variables

Matching variables	Groups			
	High–*M* ($N=25$)	Low–*M* ($N=25$)	Introversive-extratensive ($N=19$)	($N=16$)
Personal data				
Age	32.56	32.88	33.84	32.19
IQ (Wechsler-Bellevue)	116.50	113.00	114.00	120.78
Rorschach determinants				
R	24.84	19.36	22.00	22.94
M	3.84*	.60*	4.05*	1.13*
FM	4.16	3.04	4.11	3.19
m	.60	.52	.68	.50
C'	.80	1.16	.53	1.56
c	1.80	.84	1.32	.94
F	10.08	8.76	9.05	9.81
F + % (Beck)	83.88	80.60	85.53	82.13
FC	1.72	1.88	1.20*	2.31*
CF	1.04	1.88	.42*	2.81*
C	.48	.32	.08*	.94*
Sum C	2.60	3.22	.92*	5.34*

* Differences between the means of High- and Low-M groups in M and between means of introversive and extratensive groups in M, FC, CF, C, and Sum C are significant beyond the .05 level.

---

[2] It should be noted that this study is limited to the role of the human movement determinant with all other movement percepts, animal and inanimate movement (*FM* and *m*), controlled. Previous studies [9, 18, 22] have yielded mixed results, occasionally suggesting that *FM*, *m*, and *M* have common behavioral correlates, occasionally suggesting differences between them. Since Rorschach's original formulations were clearly related to the experience of *human* movement, the *M* responses have been singled out in this study for the sake of simplicity in design.

[3] Diagnoses were made by admitting psychiatrists and were corroborated by ward psychiatrists. There seemed to be little relationship between the subtype of schizophrenia and the Rorschach protocols of the 200 patients from whom the experimental samples were drawn. Most patients revealed paranoid delusions after some probing in any case, and patients with rather similar delusional content could be found in both High- and Low-*M* groups.

2. Motor inhibition task
3. Brief interview

The waiting-room observation was a procedure modified from studies carried out by Rickers-Ovsiankina [14]. The *S* was conducted into the ward visitors' room, a customary waiting room in the hospital. He was informed that *E* would call for him in about 15 minutes and was told to spend his time as he chose. The *S* was seated near a table on which were spread a large number of puzzles and colorful test objects, including such things as Isochromatic Color Plates, Lowenfeld Mosaic pieces, various mechanical aptitude tests such as nut and bolt assembly, etc. At the far end of the room, apparently working on some papers, was the observer, who was almost always unknown to the patient. The observer recorded, according to prearranged plan, all movements, activity, and speech of the patient during the 15-minute period. Recording sheets broken down into 15-second intervals were used, and the observer entered each change in position, obvious shift in attention, gross movement or comments, manipulation of materials, etc. These observations proved relatively easy to make, particularly since there were no others in the room and clear vision of *S*'s activity was possible. The presence of the observer did not seem to disturb the patients, who were accustomed to finding strangers seated in this room while waiting for psychiatric consultations. Subsequent questioning revealed that only two paranoid patients guessed the function of the observer. Any comments addressed to the observer by *S* were included in the protocol as part of a response to environmental stimulation.

Protocols were rated on a six-point scale for degree of motor activity and environmental response. Each point was defined as an aid in rating based upon a preliminary sampling of the various activities manifested by patients in a waiting-room situation. The definitions were as follows:

0 = No gross motor activity other than stretching, scratching, smoking, minor finger movements, foot tapping.

1 = No gross motor activity other than playing with personal objects, e.g., glasses, shuffling cards, toying with pencils, etc.

2 = Some minimal movement around room, reading, looking at objects on table (without manipulating them), or doing crossword puzzle brought into the room by *S*.

3 = Some speech, some movement around room, evidence of hallucination, laughing, loud and persistent whistling, other gross motor activity, e.g., picking up one or two objects from table and examining them.

4 = Fairly extensive motor activity, playing with many objects on table, or a few such objects for a very long time, occasional comments to observer or to self.

5 = Considerable restless motor activity or speech, noticing and manipulating many objects, addressing observer or persons passing outside the room.

Protocols were rated by two judges without knowledge of *S*'s name or group. One judge made separate ratings a month apart as well. Rater reliability was determined by product-moment correlations. Final waiting-room data consisted of the average of the three ratings.

The motor inhibition task consisted of a procedure adapted from the Downey Will-Temperament Scale. This task called for *S* to write the phrase "New Jersey Chamber of Commerce" as slowly as he possibly could without lifting pencil point from the page or stopping motion of the pencil. This task had been employed in prior research in this series [9]. The *S*s were informed that they were helping in the standardization of a new testing procedure. The task proved relatively simple and was easily comprehended by the patients, who cooperated readily enough. As a check on cooperation a writing sample at normal tempo was obtained to be sure that *S* had made a definite effort to inhibit his motor response. The test phrase was written in pencil across an 8-inch page between ½-inch parallel lines to maintain approximate uniformity in size. Data for this task were based on the inhibition times taken in seconds.

The interview followed immediately and consisted of an exploration within the limits of the patient's accessibility, of *S*'s daydream and other thinking activity, its preferred form and content. Patients' comments concerning their own motor activity were considered as well. It was not possible at this time, however, to get very much information concerning motivational aspects of the patients' fantasy life.

The Rorschachs, administered originally by the hospital staff in the course of diagnostic examinations, were rescored by the senior author. In the few cases of discrepant scoring the judgment of a third experienced clinician was relied upon for a final decision. Klopfer's scoring was used.

Results

The human movement determinant. The relative simplicity of the motor inhibition situation resulted in good cooperation by patients.[4] Even those *S*s with lowest inhibition times wrote the test phrase at least four times as slowly as the normal writing sample. Scoring of the waiting-room protocols according to the activity scale also proved relatively simple. The average interrater reliability coefficient (based on *z* transformations) was $r = +.85$. Data

---

[4] A third experimental situation, mirror tracing, had to be abandoned early in the development of the study when it proved too difficult and disconcerting to *S*s, who frequently rejected it.

1782 4.8 0

presented for this variable are based on the average of three ratings. Results for the High- and Low-*M* groups and the introversive-extratensive group on both variables are presented in Tables 2 and 3. It should be noted that where specific directional differences were predicted the value of $p/2$ is employed to indicate significance.[5]

Table 2
Inhibition Times and Waiting-Room Activity Scores for High- and Low-*M* groups

Groups	Experimental situation means	
	Inhibition time (secs.)	Activity rating
High-*M*	277.56**	1.84**
Low-*M*	138.40	3.09

** Difference between group means is sig. at $p < .01$.

Table 3
Inhibition Times and Waiting-Room Activity Ratings for Introversive and Extratensive Groups

Group	Experimental situation means	
	Inhibition time (secs.)	Activity rating
Introversives ($M > Sum\ C$)	266.16**	1.66**
Extratensives ($Sum\ C > M$)	139.63	3.28

** Difference between group means significant at $p < .01$.

From the data it seems clear that the High-*M* group shows considerably longer inhibition time and considerably lower activity scale ratings than the Low-*M* group. These results tend to support the hypotheses. A product-moment correlation between *M* and inhibition time yields an r of .29, which is significant at $p < .05$. The product-moment r for *M* and waiting-room activity scores is $-.23$ which barely misses significance. The correlation between *M* and inhibition time is somewhat lower than that obtained in earlier studies with normal *S*s [9], but this most likely reflects the typically greater variability of schizophrenic *S*s as well as some procedural differences between the experiments.

Previous studies had suggested the value of considering certain qualitative aspects of the *M* response [18, 19]. Certainly a corollary of the major hypothesis would be that those *S*s who produce *M* responses that are more active or vigorous in the ascribed movements may be expected to show greater inhibition and less spontaneous motor activity in their own behavior. To test this possibility the human movement responses for the 25 High-*M* *S*s were rated for *active* ("Man running, girls dancing") versus *static* movement ("Woman staring, man sleeping"). It was possible then to subdivide the High-*M* group into the 15 *S*s who showed two or more active *M* and the 10 who showed less than two such responses. The results presented in Table 4 reveal that even within the High-*M* group significant differences emerge; those *S*s who emphasize active *M* tend to have longer inhibition times and lower activity scores than the patients who gave more static *M*. When the total sample of 50 *S*s was broken down into those with preponderantly active or static *M* a similar finding emerged. These rather clear-cut findings lend further support to the hypothesized relationship between motion perception and motor inhibition. They also suggest, pending further research, a necessary refinement in interpretation of Rorschach, i.e., consideration of the vigorousness of movement. Exploratory scoring of *M* as flector and extensor and aggressive ("men with clubs") or passive ("young cherubs dancing") content failed to yield any additional information relevant to the hypotheses.

The experience-type. Results in motor inhi-

Table 4
Inhibition Times and Waiting-Room Activity Ratings for High-*M* *S*s Producing Active and Static *M*

Group	Experimental situation means	
	Inhibition time (secs.)	Activity rating
Active-*M* ($N = 15$)	377.9**	1.40*
Static-*M* ($N = 10$)	127.1	2.50

* Difference between group means significant at $p < .02 > .01$.

** Difference between group means significant at $p < .01$.

---

[5] Since the contrasted groups did not differ significantly in total responses (R) the absolute number of *M* was employed. Statistical checks on the influence of *M*% and other determinants as a function of R did not alter the trend of the data.

bition and waiting-room activity for the introversive and extratensive experience-type groups are presented in Table 3. They reveal significant differences paralleling those with *M* alone. Introversive *S*s show significantly longer inhibition times and significantly less waiting-room activity than do extratensives. Since these results, obtained with groups differing in both *M* and *C*, are quite similar to those obtained when only the number of *M* was used in separating groups, they suggest that it is the *M* response alone rather than the *M:Sum C* ratio which is crucial in relation to motor activity. It should be kept in mind that several *S*s in the High-*M* group were technically extratensive; they showed even more color responses than human movement responses. This dominant role, played by the *M* response alone, while in accord with the sensory-tonic theory casts some doubt upon Rorschach's emphasis on the experience-type per se.

Closer scrutiny of the results, however, does indicate the possibility that the color responses in the ratio do have some influence, particularly in the waiting-room situation, which may evoke spontaneous motor response. Thus, inspection of Table 2 indicates that when *M* alone is the basis for group division the difference in waiting-room scores for the contrasting groups is somewhat below that which obtains when both *M* and *C* are considered as in Table 3. In the breakdown by active-static *M* in Table 4, the waiting-room score differences between the groups are again lower, although significant. In both High- and Low-*M* and introversive-extratensive analyses, partialling out of the color factor reveals a slight although insignificant increase in the negative correlation of *M* and waiting-room activity.

Far more suggestive of a real influence of color, however, are the results obtained when the actual ratio of *M:Sum C* rather than introversive-extratensive types are considered. Statistical treatment of this ratio involved division of the High- and Low-*M* *S*s into High- and Low-*C* subgroups using a *Sum C* of 3.0 (the sample mean) or less as a basis for assignment to the Low-*C* group. Results are presented in Table 5, which reveals that motor inhibition scores vary chiefly as a function of the number of *M*. On the other hand, a marked increase in waiting-room activity scores is observable as we move from the High-*M*:Low-*C* group through High-*M*:High-*C*, Low-*M*:Low-*C*, and, finally, Low-*M*:High-*C*. This is so despite the fact that the High-*M*:High-*C* group actually averaged more *M* responses than the High-*M*:Low-*C* group. Although the differences from subgroup to subgroup are not significant, the regularity of the increase in activity with the increasing weight of color in the experience-type ratio points to a clear influence of the color factor upon the waiting-room behavior. Analysis of variance for the waiting-room activity scores in the four subgroups yields an *F* of 4.36 which is significant beyond the .01 level.

Table 5

Inhibition Times and Waiting-Room Activity Ratings for High- and Low-*M* *S*s Subdivided by *Sum C* Scores

Group	Rorschach means		Experimental situation means	
	M	*Sum C*	Inhibition time (secs.)	Activity rating
High *M*: Low *C* ($N=17$)	3.1	0.8	247.3	1.66
High *M*: High *C* ($N=8$)	5.1	6.3	298.0	2.23
Low *M*: Low *C* ($N=13$)	0.4	1.4	139.4	2.82
Low *M*: High *C* ($N=12$)	0.8	5.0	127.0	3.34

Qualitative observations. Although each patient was interviewed briefly following the motor inhibition task, the severity of the pathology in many cases prevented satisfactory communication. To some extent those individuals who reported tendencies toward active fantasy lives such as heroic daydreams or reviewing of past events or future plans tended to fall in the High-*M* group. Some *S*s who showed Low-*M* production but comparatively little motor activity reported that they tended to lose themselves in viewing television or in focusing on minute bodily changes with consequent somatic delusions.

On the whole, patients in the Low-*M*:High-*C* category tended to have spent more time on the "violent" ward because of hyperactive behavior. High-*M*:Low-*C* patients were more likely to be found on the privileged ward or were considered relatively "well-behaved," albeit severely ill, by ward personnel. In rough confirmation of Rorschach's [15] discussion of persons with "dilated" experience-types, it was observed that High-*M*:High-*C* *S*s tended to be among the most "interesting" patients in the hospital. They were usually better known to personnel and tended to demand and receive more attention while "coarctated" Low-*M*: Low-*C* *S*s, for example, tended to be less known by name and more inconspicuous.

An example of a patient whose record showed a High-*M*: High *Sum C* ratio and who typified characteristics of a "dilated" experience-type may be cited. This man was a severely disturbed paranoid schizophrenic whose ratio was 8:24, revealing more *M* and more *C* than any other patient in the population of 200 from which the equated groups were drawn. Although stemming from a low socioeconomic background this man attracted considerable attention because of the variety of his talents, fluency and eloquence of his speech, and his physical activity and vigor. On the one hand he revealed superior linguistic gifts and artistic talent while at the same time he was the outstanding athlete in the hospital. This man's drive for physical activity was manifested in his performing calisthenics between innings of pitching softball. He spoke rapidly, loudly, and almost interminably in group therapy or in conversation. Yet he was capable of sitting on the ward for long periods of time, almost completely motionless, weaving elaborate and highly imaginative fantasies of heroic achievements or of the plots devised against him. Both the *M* and *C* aspects of his personality could be readily observed although the hyperactivity predominated in behavior as in the experience-type ratio.

Discussion

The results of this study tend in general to support the experimental hypotheses. In doing so they suggest several implications that merit consideration theoretically as well as in clinical practice.

1. These findings support those of previous studies [9, 18, 22] in suggesting some validity for Rorschach's empirical association of the perception of humans in motion on the inkblots and tendencies toward inhibited or controlled motor response. They offer some evidence for the generality of the relationship inasmuch as they were obtained from a markedly pathological group of *S*s, schizophrenic patients, while the previous studies involved normal adults [9, 18] or mentally defective children [22]. Thus, despite the anticipated variability of the patients and the consequently lower correlation between *M* and the motor inhibition task, the trend for the production of several Rorschach *M* to be associated with comparatively greater capacity to inhibit writing speed emerges again. Although the low correlations preclude prediction in the individual case they suggest that the *M* response has a common interpretive significance across normal and abnormal groups. This does not deny the necessity for considering other factors in the Rorschach protocol which may counterbalance the *M* or alter its interpretation somewhat. At the same time it seems clear that as a perceptual or apperceptual phenomenon presumably exemplifying some general behavior principle, the human movement response to the Rorschach inkblots has definite correlates in behavior quite apart from the immediate testing situation.

2. Of particular theoretical interest is the nature of these correlates of *M*. The fact that individuals chosen on the basis of contrasting tendencies to perceive humans in motion on inkblots tend to differ in their capacity to inhibit motion or to move about in a waiting room supports the view that perceptual response reflects pervasive behavior characteristics. More specifically, theories implying a vicarious relationship between motion perception and overt movement, of which the sensory-tonic theory [21, 23] is an example, gain support from these findings. Whether Rorschach's *M* reflects imaginativeness or fantasy processes as well as tendencies toward motor restraint remains a question for further research. The positive results of this study offer encouragement for an attempt to explore the interrelations of motion perception, fantasy, and motor activity with the ultimate aim of developing the unified theory of thinking now only speculatively suggested by the writers cited in the introduction [8, 10, 11, 13].

3. The data of the experiment throw comparatively little light on the knotty problem of the experience-type ratio, however. Most of the data suggest that the quantity of *M* rather than the experience-type as given in $M > C$ vs. $C > M$ is most potent in determining mo-

tor responsiveness. The results presented in Table 5 suggest, however, that the role of color cannot be ignored. The amount of spontaneous motor activity in the waiting-room situation varied with a quantitative relationship between *M* and *C* somewhat different from that implied by introversive-extratensive ratios. The observed relationship suggests that while *M* is the primary factor related to motor inhibition, color, reflecting the individual responsiveness to the environment, to some extent counterbalances this tendency. Since the influence of color emerged only in the waiting-room situation (which might be considered evocative of more spontaneous performance), the interpretation of the color factor as reflecting environmental responsiveness or outgoing tendencies [6, 12, 16] would appear consistent with the data. Whether it is the color qua color or merely a tendency to respond in diffuse associations with lessened consideration of blot contour [17] which reflects *S*'s characteristic emotional or motor lability is a question beyond the scope of this investigation. Nevertheless, the finding that some influence of the actual experience-type ratio exists is quite important for validation of Rorschach's original interpretations and merits more intensive scrutiny.

A final word may be said concerning the implications of the results for research in the field of projective techniques. It seems clear that experimentation in the theory and process of clinical instruments is possible. The question seems to be one of finding concrete situations in which the somewhat nebulous personality descriptions can be expressed operationally. The interaction of various factors and the necessarily ideographic approach of the clinician need not constitute an insurmountable obstacle. The original assignment of interpretative meaning to the determinants on Rorschach's part was, as he makes quite clear, drawn from his observation of their correlates in overt behavior. Even in "global" interpretation some aspect of the "nuclear definition" [12] of the factor is necessarily preserved. Thus, one need not expect to find very high correlations in group data, but low, significant correlations should emerge, as they have in the present series of studies.

Summary and Conclusions

This study was part of a program of research designed to integrate key Rorschach determinants within a more general theory of perception. Previous studies have revealed a relationship between inhibition of motor response and the tendency to produce human movement percepts (*M*) on the Rorschach, a finding in keeping with Werner's sensory-tonic theory of perception [21] as well as with certain theories of thought and its development [4, 8, 10, 11, 13]. The present study extended this research to include a pathological group (schizophrenics) and also focused attention on the experience-type (*M:Sum Color*) ratio as well as *M* alone. It was hypothesized that schizophrenics producing several *M* percepts would show greater tendency to inhibit overt motor response than a comparable group which produced few *M*. Similarly schizophrenics with introversive experience-types were expected to show more motor inhibition and less spontaneous gross motor activity during a waiting period than extratensive *S*s. It was anticipated that the number of *M* responses would prove most effective as an index of motor inhibition, with *C* responses, however, influencing the tendency towards spontaneous motor activity.

Fifty hospitalized male schizophrenics were divided into two groups on the basis of production of two or more, or one or less, Rorschach *M* responses. These groups did not differ significantly in age, intelligence, type of schizophrenia, length of hospitalization, or in production of various Rorschach determinants other than *M*. Extreme introversive and extratensive groups were also employed equated for all factors except the *M:Sum C* ratio.

The *S*s were studied in two situations, motor inhibition and spontaneous motor activity. The former called for the writing of a brief phrase as slowly as possible without stopping motion of the pencil. The times taken to write the phrase furnished the data for this task. The second situation involved recordings by an observer of a patient's waiting-room behavior during a 15-minute period. Protocols were rated independently according to a scale of degree of gross motor activity and response to environmental objects. Interrater reliabilities yielded an average r of .85.

Results generally tended to support the hypotheses. Thus, the High-*M* group showed significantly longer inhibition time and significantly lower waiting-room activity scores than did the Low-*M* group. The introversive *S*s differed in parallel fashion from the extratensives. When the human movement percepts were rated for relative vigorousness of activity, *S*s who showed predominantly *active M* proved to have longer inhibition time and less waiting-room activity than *S*s producing *static M*. To check the role of color responses High- and Low-*M* groups were subdivided into High- and Low-*Sum C* groups. Although color did not appear materially related to motor inhibition performance, a regular increase in waiting-room activity scores emerged with the increased weight of the color in the *M:Sum C* ratio. The results of the investigation corroborate in a limited way some of Rorschach's original formulations and suggest that the interrelations of motor activity, perception of motion, and fantasy behavior offer a fruitful area for experimental research.

Received May 27, 1953.

References

1. Beck, S. J. *Rorschach's test.* Vol. II. *A variety of personality pictures.* New York: Grune & Stratton, 1945.
2. Fenichel, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.
3. Ferenczi, S. Thinking and muscle innervation. In *Further contributions to the theory and techinique of psycho-analysis.* London: Hogarth, 1950.
4. Freud, S. Formulations regarding the two principles in mental functioning. In *Collected papers.* Vol. IV. London: Hogarth Press, 1946. Pp. 13-21.
5. Goldman, A. The effect of motor activity on the autokinetic phenomenon. Paper read at East. Psychol. Ass., Atlantic City, March, 1952.
6. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.
7. Levi, J., & Kraemer, Doris. Significance of a preponderance of human movement responses in children below age ten. *J. proj. Tech.*, 1952, **16**, 361-365.
8. Lewin, K. *A dynamic theory of personality.* New York: McGraw Hill, 1935.
9. Meltzoff, J., Singer, J. L., & Korchin, S. J. Motor inhibition and Rorschach movement responses: A test of the sensory-tonic theory. *J. Pers.*, in press.
10. Murphy, G. *Personality.* New York: Harper, 1947.
11. Piaget, J. *The language and thought of the child.* New York: Harcourt-Brace, 1932.
12. Piotrowski, Z. A. A Rorschach compendium. In J. A. Brussel, K. S. Hitch, & Z. A. Piotrowski (Eds.), *A Rorschach training manual.* Utica, N. Y.: State Hospitals Press, 1950.
13. Rapaport, D. (Ed.) *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951.
14. Rickers-Ovsiankina, Maria. Studies of the personality structure of schizophrenic individuals. I. *J. gen. Psychol.*, 1937, **16**, 153-178.
15. Rorschach, H. *Psychodiagnostics.* New York: Grune & Stratton, 1942.
16. Schachtel, E. G. Projection and its relation to character attitudes and creativity in the kinesthetic responses. *Psychiatry*, 1950, **13**, 69-100.
17. Siipola, E. The influence of color on reactions to inkblots. *J. Pers.*, 1950, **18**, 358-382.
18. Singer, J. L., Meltzoff, J., & Goldman, G. D. Rorschach movement responses following motor inhibition and hyperactivity. *J. consult. Psychol.*, 1952, **16**, 359-364.
19. Thetford, W. N. Fantasy perceptions in the personality development of normal and deviant children. *Amer. J. Orthopsychiat.*, 1952, **22**, 542-550.
20. Wiezsäcker, V. v. Der Gestaltkreis. *Pflüg. Arch. ges. Physiol.*, 1933, **231**, 630-661.
21. Werner, H. Motion and motion perception: a study in vicarious functioning. *J. Psychol.*, 1945, **19**, 317-327.
22. Werner, H. Perceptual behavior of brain-injured mentally-deficient children. *Genet. Psychol. Monogr.*, 1945, **31**, 51-110.
23. Werner, H., & Wapner, S. Toward a general theory of perception *Psychol. Rev.*, 1952, **59**, 324-338.

Journal of Consulting Psychology
Vol. 18, No. 1, 1954

A Comparison of Raven's Progressive Matrices (1938) with a Short Form of the Wechsler-Bellevue[1]

Bert Levine and Ira Iscoe
University of Texas

The Progressive Matrices (1938), and a short form of the Wechsler-Bellevue Form I consisting of the Comprehension, Vocabulary, and Block Design subtests, were given to 60 undergraduates. There were equal numbers of males and females. The average age was 19.1 years. Each *S* was first given the P.M. and then the W-B subtests. Administration and scoring were in accordance with the directions given in the respective manuals. The mean score on the P.M. was 50.8 with an *SD* of 6.6. No significant sex differences were found. A Pearson *r* of .55 was obtained between the total P.M. score and the total weighted score of the three subtests. Individual weighted subtest scores correlated with the P.M. total, as follows: Comprehension .21, Vocabulary .48, and Block Design .63. Set E of the P.M., composed of the 12 most difficult items, correlated .52 with the short form of the W-B, and .23 with the Comprehension, .47 with the Vocabulary, and .56 with the Block Design subtest weighted scores. Set E by itself correlated .85 with the P.M. total. The P.M. appears to tap areas of intelligence most closely related to Block Design and not significantly related to Comprehension. As one moves away from cultural loadings on the W-B subtests, the correlation increases markedly. Inasmuch as the P.M. is easily administered, untimed, and demands no speech and only a minimum of motor ability, its possible utility as an intelligence test with people having certain types of sensory and motor handicaps as well as illiterates and foreign born deserves further investigation.

Brief Report
Received December 2, 1953

---

[1] An extended report of this study may be obtained without charge from Ira Iscoe, Department of Psychology, University of Texas, Austin, Texas, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4159 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 18, No. 1, 1954

Rorschach Performance in Relation to Two Types of Stress[1]

Murray S. Stopol
Teachers College, Columbia University[2]

In the past several years investigators have become more cognizant of individual differences observed in performance under stress. The fact that some persons show a decrement in performance, while others show improvement, and still others have an initial decrement followed by improvement, has led psychologists to attribute these seemingly inconsistent results to such factors as (*a*) the kind of stress situation employed, (*b*) the type of motivation the subject uses, i.e., ego-defensive or need-persistent, and (*c*) the subject's other personality characteristics. The problem of individual reactions to stress is of practical as well as theoretical importance. For example, if we knew more about the nature of these individual differences, it might enable us to predict who would be adversely affected by a stress situation. At the same time, selection procedures might be readily improved in schools, in industry, and especially in military service, where persons are frequently required to perform intellectual and skilled tasks under conditions of distraction or threat.

The most widely used instrument to assess personality variables associated with behavior under stress has been the Rorschach test. Conflicting evidence seems to exist with respect to the relationship between behavior under various types of stress conditions and alleged Rorschach indices of stress tolerance. Williams [19], Waldman [15, 16], Smith and George [13], and Baker and Harris [2] found a definite relationship between Rorschach measures and reactions to stress, while Allee [1], Eriksen, Lazarus, and Strange [9], Deese and Lazarus [6], and Carlson and Lazarus [4] found no relationship. Among the major possibilities that exist in attempting to resolve these differences in results are: (*a*) different aspects of behavior are being measured in these experiments, or (*b*) the validity of the Rorschach indices may be questionable, or both.

Lazarus, Deese, and Osler [11] and Deese and Lazarus [7] have presented a summary of the literature on the effects of psychological stress upon performance, including personality correlates of behavior under stress. Both reviews present numerous studies of performance under stress in which the Rorschach test, as well as other measures of personality, have been utilized. The authors conclude that the majority of these investigations report either negative findings or else present correlations which are statistically significant but are too small to be of any practical value.

Purpose and Hypotheses

The purpose of this study was to investigate the relationship between certain Rorschach measures, presumably concerned with intellectual and emotional control, and performance under two types of stress situations designated as failure stress and task-induced stress (experimental distraction). It was an attempt to examine experimentally the validity of the claim made by Rorschach workers, on the basis of clinical experience and experimentation, that certain Rorschach variables indicate how vulnerable an individual is to stress.

In a previous study, Stopol[3] found that resistance to failure stress was independent of task-induced stress tolerance. Moreover, a cumulative effect of stress was reported from one session to the next, regardless of the type used. Furthermore, no difference was found between the effects of failure stress and experimental distraction if the temporal sequence of presentation of the type of stress was controlled. Consequently, the Rorschach hypotheses under consideration were applied to each type of stress situation separately.

---

[1] The present paper is part of a dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Teachers College, Columbia University, 1951. The writer wishes to express his thanks to Drs. Edward Joseph Shoben, Jr. (doctoral committee chairman), Paul E. Eiserer, Lincoln E. Moses, and Laurance F. Shaffer for their invaluable advice and criticism.

[2] Now at Veterans Administration Hospital, Minneapolis, Minnesota and the University of Minnesota.

[3] Stopol, M. S. The consistency of stress tolerance. To be published.

General hypothesis. The Rorschach test is able to differentiate individuals more able to tolerate and individuals less able to tolerate (*a*) experimentally induced failure stress and (*b*) task-induced stress (experimental distraction).

Specific hypotheses.[4] The following specific hypotheses were tested:

1. The greater the amount of Z, the greater the tolerance for stress.
2. Individuals with R between 25 and 50 will be more able to tolerate stress than individuals outside that range.
3. Individuals with $F\%$ between 25 and 50 will be more able to tolerate stress than individuals outside that range.
4. Individuals with $F+\%$ between 69 and 89 will be more able to tolerate stress than individuals outside that range.
5. The higher the $F+\%$ on the color cards, the greater the tolerance for stress.
6. The greater the preponderance of FC over $CF + C$, the greater the tolerance for stress.
7. The greater the preponderance of form-dominant responses over form-subordinate responses, the greater the tolerance for stress.
8. The fewer the number of color shock signs, the greater the tolerance for stress.
9. Individuals with $Av.\ T/1R$ between 9 sec. and 16 sec. will be more able to tolerate stress than individuals outside that range.
10. The higher the mean rank based on a combination of low $W\%$, high $Dd\%$, and the number of form-dominant responses greater than form-subordinate responses, the greater the tolerance for stress.
11. Individuals who show a coarctative or extratensive pattern will perform less well under stress than individuals who show an introversive or ambiequal pattern.
12. The greater the number of Davidson's [5] signs of effective adjustment, the greater the tolerance for stress.

The specific hypotheses used in this study stem from three sources. The variables used in hypotheses 1, 2, and 9 were suggested by Williams [19]. Hypotheses 3, 4, 6, 10, and 11 were largely based on the results of previous investigators [2, 10, 13, 15, 16, 19] who found one or another of the particular measures in question significantly related to performance under various types of stress conditions. Hypotheses 5, 7, 8, and 12, on the other hand, were derived from clinical experience. In general, the Rorschach factors under investigation are supposedly related to intellectual and emotional control, the necessary characteristics presumably involved in the adequate functioning of an individual in a stress situation.

Method

A detailed description of the method is given in a previous paper[5] and the original report [14]. The Rorschach test, administered according to the directions outlined by Beck [3], and a series of five Wechsler-Bellevue Form II Digit-Symbol Test (DST) [17] practice trials were administered to 48 male Columbia University students during the first session of the experiment. The second session, which took place within 24 to 48 hours after the first, consisted of two periods of three phases each. During the first period in Phase *a*, each *S* was given 10 DST trials under optimal conditions. In Phase *b*, either failure stress or experimental distraction was administered during five of the 10 trials. Phase *c* was a poststress phase where no stress was induced during a block of eight trials. The second period began after a 15-minute interval. Each *S* underwent a procedure similar to the first period except that a different form of stress was induced. Two types of stress conditions were utilized: (*a*) failure stress, which consisted of a series of admonishing and reproving comments strongly implying disapproval of performance, and (*b*) task-induced stress or experimental distraction, which involved a series of bells, a buzzer, and a flashing light which were presented during a given DST trial.

The stress score utilized was the degree of change in performance on the DST during the stress trials as compared to prestress level of functioning within each period. This score was calculated from the mean of the last five prestress trials minus the mean of the five stress trials, divided by the range of the last five prestress trials.

---

[4] A detailed description of the criteria used in scoring the various Rorschach measures under consideration may be found in the original report [14].

[5] See footnote 3.

Results

Before reporting the actual results, a word might be in order concerning the statistical techniques used in testing the hypotheses. The method of rank-order correlation was used in testing hypotheses 1, 5, 6, 7, and 10. The chi-square method was applied in the remaining hypotheses. For hypotheses 8 and 12, chi square was used rather than rank-order correlation because of the narrow range of scores as well as the presence of many ties. The cut-off point was made as near the median as possible. A single-tailed test of significance was used in all instances.

In applying the chi-square test to the data in this study, it was noted and anticipated that the expected cell frequencies would be rather small. It was therefore planned beforehand to calculate the p values of chi square by Fisher's exact method if an obtained value of chi square, corrected for continuity, was in the predicted direction and had a p value of .10 or less. This is in line with Edwards' [8] suggestion.

To check on the scoring reliability of the various Rorschach factors used in this study, every third protocol was given to a qualified Rorschach examiner for independent scoring.[6] The total number of responses scored was 677, involving 779 determinants. Both main and additional responses were tabulated. The scoring reliability for all measures was extremely high, no percentage of agreement falling below 90. Thus we may readily consider the scoring of the Rorschach protocols in this study to be highly reliable.

For purposes of analysis, *S*s having task-induced stress initially, followed by failure stress will be referred to as Group A. The *S*s having failure stress initially, followed by task-induced stress will be called Group B. Since Group A was found to be significantly different from Group B as a result of the cumulative effect of stress, regardless of the type used,[7] each group was necessarily treated separately.

For Group A, no significant relationship was found between any of the hypothesized Rorschach variables and resistance to either failure stress or experimental distraction.

For Group B, the only significant relationship found was between the combination of low $W\%$, high $Dd\%$, and a greater number of form-dominant than form-subordinate responses and the ability to withstand both failure stress and experimental distraction. The rank-order correlation coefficients were .37 and .47 respectively. Both coefficients are significantly different from zero beyond the .05 level using a one-tailed test. This Rorschach measure was taken literally from Waldman [15, 16] who found this combination to differentiate "slow" from "rapid" learners under stress. The score was obtained by taking the mean rank of each person on each factor in the combination. Thus, an individual who was ranked low on two factors and extremely high in another may be above the median when these three factors are averaged. It is, therefore, somewhat misleading, if not hazardous, to generalize about a combination of *low* $W\%$, *high* $Dd\%$, and a *greater* number of form-dominant over form-subordinate responses. The legitimacy or justification of using such a measure may thus be seriously questioned. In examining the data further, if we classify each person on the basis of being above or below the median value in $W\%$, $Dd\%$, and form-dominant greater than form-subordinate responses, in Group B seven out of 24 *S*s showed all three signs in the predicted direction; in Group A only six *S*s presented all three signs in combination. Of the six *S*s in Group A, none was above the median value for task-induced stress tolerance; three cases were above the median value for failure stress tolerance. Of the seven *S*s in Group B, four were above the median value for both failure stress and task-induced stress tolerance. It appears obvious that this hypothesis fails to hold when subjected to a more stringent and meaningful test. Because of the vagueness of the proposed criteria, and since this hypothesis was not tenable for Group

---

[6] The writer would like to thank Dr. Shabse H. Kurland, who so generously consented to independently score the Rorschach protocols. Prior to giving the protocols to Dr. Kurland, various scoring samples from both Beck [3] and the remaining 32 records were discussed in detail.

[7] See footnote 3.

A, it is highly possible that the significant correlations obtained with Group B may reflect only chance factors. Furthermore, this suggestion is reinforced when we consider that the probability of obtaining two significant correlations at the .05 level by chance when 24 tests of significance are calculated is .34 [18]. Thus we may conclude that the Rorschach test as used in this study is of no value in predicting efficiency of performance under conditions of failure and experimental distraction.

Discussion

Only one Rorschach measure, a combination of low *W*%, high *Dd*%, and a greater number of form-dominant over form-subordinate responses, was found to be associated with the ability to withstand both failure stress and experimental distraction if failure stress was followed by experimental distraction. The difficulty of testing this measure and the vagueness of this combination were noted. A more appropriate statistical analysis found these factors to have no diagnostic utility, since they could have resulted from chance. Therefore, for all practical purposes, it may be said that no Rorschach modality was consistently found to be associated with behavior under either type of stress situation.

Failure to obtain significant positive results with those Rorschach factors allegedly related to stress tolerance is open to several interpretations. One possibility is that these particular Rorschach measures possess no utility as predictors of failure stress or task-induced stress tolerance. The previous investigators who did find significant relationships may simply have been capitalizing on chance factors. It may also be possible that the Rorschach is able to differentiate between persons with respect to tolerance for a given stress, but by measures other than those hypothesized in this investigation. Another possibility may be that these Rorschach signs are not sensitive enough to discriminate between individuals subjected to an experimentally induced stress which is probably not as severe or complex as some extra-laboratory stresses of a longer duration. Still another possibility may be that the results are a function of the *S*s investigated. It should be further noted that the kinds of behavior which these Rorschach factors are attempting to predict may not be sufficiently unitary to be predictable by an instrument such as the Rorschach, and especially by the so-called "sign approach." If we may speculate, the latter possibility appears to be the most feasible of those suggested. It has been clearly demonstrated that a person's tolerance for stress, to a large extent, depends upon both the quality of the particular stressful stimulus condition to which he is exposed[8] and the type of motivation mobilized in a particular type of stress situation [7, 11]. Furthermore, Sherman and Bell [12] suggest that tolerance for a given stress may be a function of many conditions. If this is the case, stress tolerance may be regarded as a complex phenomenon, nonunitary in nature. Moreover, prediction of such behavior seems almost impossible in terms of a one-to-one relationship with other variables, as was the situation in this investigation. Nevertheless, having no logical basis for selecting any of these possibilities as holding true, we must regard these results as indicating that a number of Rorschach signs presumably associated with resistance to stress fail to meet this particular test of experimental validation.

Summary and Conclusions

The purpose of this study was to determine whether certain measures derived from the Rorschach test were able to predict the ability of persons to withstand two distinct types of stress conditions designated as failure stress and task-induced stress (experimental distraction). A series of 12 hypotheses was evolved.

The Rorschach test and a series of Digit-Symbol Test trials were given to a group of 48 *S*s in two sessions. The second session consisted of two periods, each involving a prestress, a stress, and a poststress phase. Twenty-four *S*s received failure stress initially, followed by task-induced stress, while 24 *S*s experienced task-induced stress initially, followed by failure stress.

Except for a combination of low *W*%, high *Dd*%, and a greater number of form-dominant over form-subordinate responses, which was correlated with the ability to withstand both failure stress and experimental distraction if failure stress was followed by experimental distraction, no specific Rorschach measure was found to be significantly related to the ability to resist either type of stress situation. It was shown that the one significant relationship probably reflected chance factors, since the probability of obtaining two statistics significant at the .05 level by chance when 24 hypotheses are tested is .34. It may therefore be concluded that the Rorschach factors employed in this study are of no value in predicting efficiency of behavior under conditions of failure stress or task-induced stress.

---

[8] See footnote 3.

Received May 29, 1953.

References

1. Allee, Ruth. Rorschach responses of extreme deviation in an experimental stress condition. Unpublished master's thesis, Univer. of Missouri, 1948. (Cited in F. McKinney, Ruth R. Hines, G. B. Strother, & Ruth Allee. Experimental frustration in a group test situation. *J. abnorm. soc. Psychol.*, 1951, **46**, 316-323.)
2. Baker, L. H., & Harris, Jane S. The validation of Rorschach test results against laboratory behavior. *J. clin. Psychol.*, 1949, **5**, 161-164.
3. Beck, S. J. *Rorschach's test.* Vol. I. *Basic processes.* New York: Grune & Stratton, 1950.
4. Carlson, V. R., & Lazarus, R. S. A repetition of Meyer Williams' study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1953, **17**, 247-253.
5. Davidson, Helen. A measure of adjustment obtained from the Rorschach protocol. *J. proj. Tech.*, 1950, **14**, 31-38.
6. Deese, J., & Lazarus, R. S. The effects of psychological stress upon perceptual-motor performance. USAF *Hum. Resour. Res. Cent., Res. Bull.* 1952, No. 52-19, 1-15.
7. Deese, J., & Lazarus, R. S. The effects of psychological stress upon performance: a theoretical analysis and review of recent developments. Dittoed manuscript, Sept., 1953.
8. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
9. Eriksen, C. W., Lazarus, R. S., & Strange, J. R. Stress and its personality correlates. *J. Pers.*, 1952, **20**, 277-286.
10. Hertzman, M., Orlansky, J., & Seitz, C. P. Personality organization and anoxia tolerance. *Psychosom. Med.*, 1944, **6**, 317-331.
11. Lazarus, R. S., Deese, J., & Osler, Sonia J. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293-317.
12. Sherman, M., & Bell, Elizabeth. The measurement of frustration; an experiment in group frustration. *Personality*, 1951, **1**, 44-53.
13. Smith, S., & George, C. E. Rorschach factors related to experimental stress. *J. consult. Psychol.*, 1951, **15**, 190-195.
14. Stopol, M. S. An experimental investigation of the consistency of stress tolerance and related Rorschach factors. Unpublished doctor's dissertation, Teachers Coll., Columbia Univer., 1951.
15. Waldman, M. Personality factors in reaction to a stressful psychomotor task. Paper read at East Psychol. Ass., Worcester, April, 1950.
16. Waldman, M. Personality factors and performance under stress in schizophrenics. *Amer. Psychologist*, 1951, **6**, 314. (Abstract)
17. Wechsler, D. *The Wechsler-Bellevue Intelligence Scale. Form II.* New York: Psychological Corp., 1946.
18. Wilkinson, B. A statistical consideration in psychological research. *Psychol. Bull.*, 1951, **48**, 156-158.
19. Williams, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 21-29.

Journal of Consulting Psychology
Vol. 18, No. 1, 1954

Age, Sex, and the Manifest Anxiety Test[1]

A. W. Bendig
University of Pittsburgh

A shortened (50-item) form of Taylor's Manifest Anxiety Scale [1] was given to 497 students enrolled in undergraduate psychology courses. The *Ss* also recorded their sex and age on their answer sheets. Raw MAS scores were converted to normalized McCall *T* scores for the subsequent analysis. The *Ss* were divided into seven age groups and two sex groups and their scores analyzed by an analysis of variance. Mean MAS scores were also computed for each age and sex group for inspection for trends. Finally, a critical ratio was computed between the mean MAS scores of the Iowa [1] and Pittsburgh samples and an *F* ratio computed from their variances.

Analysis of variance showed no significant differences (at the .05 level) between the sex or the age groups, nor was the sex-age interaction term significant. While there was not a significant difference between the mean MAS scores of men and women subjects, women show slightly higher mean MAS scores in six of the seven age groups represented. A slight curvilinear (parabolic) relationship between MAS and age can be noted for the men with a peak occurring in the 23-25 year age group. This slight curvilinearity is not apparent for women.

A comparison of the mean transformed MAS scores for the Iowa and Pittsburgh samples gave a critical ratio of 2.73 which is significant at the .01 level. The *F* ratio between the variances of the two samples was 1.11 which is not at all significant. However, the difference between the two means was very small.

Apparently, scores on the Manifest Anxiety Scale are not related to sex differences or to age differences over the range represented by our sample of college students. Selecting students on the basis of extreme MAS scores, which is commonly done in comparing the learning rate of "anxious" and "nonanxious" subjects, does not introduce age or sex differences between the subjects comprising the two groups. The significant difference between the mean MAS scores of the Iowa and Pittsburgh samples may be attributed to population differences. Apparently college students demonstrate significantly less manifest anxiety at Pittsburgh than they do at Iowa although the magnitude of this difference is small.

Brief Report

Received December 4, 1953

Reference

1. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285-290.

---

[1] An extended report of this study may be obtained without charge from A. W. Bendig, Dept. of Psychol., Univer. of Pittsburgh, Pittsburgh 13, Pa., or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4158 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 18, No. 1, 1954

The Rorschach White Space Response and "Oppositional" Behavior[1]

Albert Bandura
State University of Iowa[2]

The purpose of the present study was to investigate the validity of some of the assumptions underlying the clinical interpretation of the Rorschach white space response (*S*). In general, despite a voluminous clinical and expositional literature, the basic principles of Rorschach interpretation have not been subjected to critical examination, and there have been relatively few attempts at systematic experimental validation [2, 15]. This is as true of the *S* response as it is of the other major scoring categories.

There is fairly extensive agreement among Rorschach workers concerning the psychological values to be assigned to the *S* response. It is assumed to reflect some sort of "oppositional" tendency on the part of the subject. Rorschach tradition holds further that there are several types of overt behavioral expression possible for this inferred oppositional tendency, the type of expression being determined by the basic personality configuration. This configuration is usually defined in terms of the introversion-extroversion continuum. If the *S* responses occur in an *extratensive* personality pattern, the opposition will be directed against the environment, resulting in behavior characterized as "assertive," "stubborn," "negativistic," and "defiant." If they occur in an *introversive* personality pattern, the opposition will be directed inwardly against the self, resulting in behavior characterized as "inadequacy feelings," "self-distrust," and "self-criticism." In an *ambiequal* personality pattern, the *S* responses indicate that the oppositional trend will be expressed in behavior characterized as "ambivalent tendencies" or "indecision."

Validation studies concerned primarily with the *S* response are for the most part lacking. A review of the available evidence on the relationship between the *S* response and oppositional behavior tendencies reveals some contradictory findings. A number of studies have reported a positive relationship between the occurrence of *S* responses and antisocial behavior [3, 12, 16]. However, these studies present some serious methodological defects, and none includes a comparison with an adequate control group. On the other hand, Endacott's [6] study in which delinquent boys were compared with several normal control groups failed to show any significant differences among the groups in the frequency of occurrence of *S* responses. Two studies specifically designed to evaluate the validity of the *S* response [7, 10] have recently been reported. The findings of both were essentially negative. However, the adequacy of some of the criterion measures used in these studies is rather questionable. In summary, the validity of current interpretations of the *S* response is still to be established.

Statement of the Problem

The present investigation is concerned with the validity of the inferences drawn from the *S* response as reflecting an "oppositional" tendency. Teachers' ratings were used as the independent criterion measure against which to evaluate the Rorschach hypotheses. The specific hypotheses subjected to empirical test were:

a. A significant and positive relationship exists between the number of *S* responses and ratings of "negativism" in extratensive subjects.

b. A significant and positive relationship exists between the number of *S* responses and ratings of "inadequacy" in introversive subjects.

---

[1] This paper is a modification of a portion of a doctoral dissertation submitted to the Department of Psychology at the State University of Iowa. The writer wishes to express grateful acknowledgment to Dr. Arthur L. Benton for his direction of this investigation and to Dr. Harold P. Bechtoldt for many helpful suggestions.

[2] Now at Stanford University.

c. There is no systematic relationship between ratings of "inadequacy" and the number of *S* responses in extratensive subjects and, similarly, there is no systematic relationship between ratings of "negativism" and the number of *S* responses in introversive subjects.

Method

Subjects. A total of 81 University of Iowa High School students, of whom 33 were females and 48 were males, served as subjects (*Ss*). They ranged in age from 13 to 18 years, the median age being 15 years.

Procedure. The *Ss* were tested individually. The Rorschach administration differed from the conventional technique only with respect to length of exposure time. While in the standard procedure *S* is usually permitted to look at the cards for as long or short a time as he desires, in this experiment each *S* was required to look at each card for a four-minute period. The Rorschach inquiry was conducted in the conventional manner, i.e., immediately after the *S* had given his response to all ten of the cards.

Rorschach scoring procedure. The Rorschach records were scored according to the scoring system of Hertz. In computing the *S* score, primary space responses were counted as 1 and secondary space responses were counted as .5. The color and movement responses were used to determine the experience types which were defined as follows:

Extratensive: $\Sigma C > M$
Ambiequal: $\Sigma C = M$
Introversive: $M > \Sigma C$

The Rorschach experience-type classification was used in checking the hypotheses concerned with the behavioral inferences drawn from *S* responses in the different experience types. Twenty-one *Ss* who fell into, or close to, the ambiequal category were excluded from the analysis. Thirty-two of the most extratensive and 28 of the most introversive *Ss* were selected for the analysis. One of the introversive *Ss*, who gave a five-response protocol and rejected six of the cards, was excluded from the final analysis. The average $\Sigma C{:}M$ value for the extratensive group was 6.2:1.6 while the average $\Sigma C{:}M$ value for the introversive group was 1.9:3.6.

Rating procedures. Five teachers rated the 59 *Ss* on four personality variables, viz., negativism, assertiveness, inadequacy feelings, and self-distrust. In terms of the statements in the Rorschach literature, the first two personality characteristics would reflect the behavioral expression of *S* responses in an extratensive pattern, whereas the last two would reflect the behavioral expression of *S* responses in an introversive pattern. Each personality variable was represented on a graphic rating scale by five successive descriptive statements indicating varying degrees of the behavior being rated. Simple descriptions of observable behavior were used, and the maximal rating was placed sometimes to the right and sometimes to the left in order to minimize a possible response tendency on the part of the rater to place *S* in the same category in each of the ratings. The number of *Ss* rated by any one teacher varied as follows: 57, 47, 40, 33, and 11. Fifty-four of the 59 *Ss* were rated by at least three or four teachers. The remaining five *Ss* received two ratings.

The rating procedure consisted of a modified rank-order method. Each rater was presented with a pile of cards containing the names of the *Ss* he could rate. The first step was to sort these cards into five consecutive piles, each pile describing a degree of the behavior being rated. Having done this, the teacher then ranked the names within the piles thus providing a final rank order for the group. A rank of 1 was regarded as the highest in the series, i.e., was assigned to the *S* who was considered most negativistic, most assertive, most inadequate, or most unsure of himself.

The assigned ranks were converted into scale values on the base line of the normal curve by Hull's method [9]. The numerical values obtained from the rank data constituted the operational definition of these personality variables. The mean value for each *S* on each of the personality variables was computed to give a final composite score which was used to evaluate the Rorschach hypotheses.

Results

Reliability of the S *score and ratings.* Since the distribution of *S* responses over the first five cards (51%) was found to be comparable to the distribution over the last five cards

(49%), the split-half technique was employed in estimating the reliability of the *S* response. The correlation between the two halves was found to be .72. The estimate of reliability of the whole test, as determined by the use of the Spearman-Brown formula, was .84. The magnitude of this correlation shows the *S* response to be a consistent response tendency.

The reliability of the ratings was estimated by the use of an analysis of variance technique for incomplete sets of ratings, as described by Ebel [4]. The size of the obtained reliability coefficients (negativism, $r = .83$; assertiveness, $r = .79$; inadequacy feelings, $r = .71$; self-distrust, $r = .81$) reflects considerable consistency among the teachers' ratings.

Intercorrelation of the ratings. Of the four personality variables rated, negativism and assertiveness were designed to reflect the behavioral expression of *S* responses in an extratensive pattern while the inadequacy and self-distrust ratings were designed to reflect the behavioral expression of *S* in the introversive pattern, as assumed in the Rorschach hypothesis. The relationship between these personality characteristics, as measured in this study, was estimated by means of product-moment correlations.

A fairly high and positive relationship ($r = + .64$) was obtained between ratings of inadequacy and self-distrust, which suggests that these two personality characteristics are closely related. A less definite relationship is revealed in the correlation between ratings of negativism and assertiveness ($r = + .32$). There were inverse relationships between ratings of assertiveness and self-distrust ($r = - .74$) and between assertiveness and inadequacy feelings ($r = - .45$). The ratings of negativism seem to be relatively independent of inadequacy feelings ($r = + .10$) or self-distrust ($r = + .04$). In general, the size of the intercorrelations among the ratings is contraindicative of the existence of much "halo" effect.

Relationship of S *responses to personality ratings.* In testing the hypotheses with respect to the behavioral inferences drawn from the *S* response, product-moment correlations were calculated between the *S* scores and the average ratings of negativism, assertiveness, self-distrust, and inadequacy feelings for the two Rorschach experience types. Since the distribution of *S* scores departed from normality in the direction of positive skewness, the $\sqrt{X + .5}$ transformation was employed and the correlations were run on the transformed data. The results of this analysis are shown in Table 1.

The negativism ratings and the *S* scores in the extratensive and introversive groups correlated + .27 and + .44 respectively. Assertiveness was found to be correlated + .26 with *S* scores in the extratensive group and + .13 in the introversive group. No relationship was revealed in the correlation of + .02 between inadequacy feelings and the *S* scores in the extratensive group and + .09 between these measures in the introversive group. An inverse relationship was found between ratings of self-distrust and *S* scores in both the extratensive and the introversive group (−.23) and the introversive group (− .30).

A Z' transformation was used to test the significance of the differences between the correlation coefficients in the two experience types for the different behavior ratings. In no case did the values of r differ significantly between the two experience types. Since the hy-

Table 1

Table of Correlations between *S* Scores and the Personality Ratings for the Total Sample and for Two Rorschach Experience Types

Ratings	Experience type		Weighted average ($N=59$)	*RS* ($N=59$)	Partial correlations ($N=59$)
	Extratensive ($N=32$)	Introversive ($N=27$)			
Negativism	+.27	+.44	+.35	+.15	+.34**
Assertiveness	+.26	+.13	+.20	+.15	+.14
Inadequacy	+.02	+.09	+.05	−.12	+.18
Self-distrust	−.23	−.30	−.26	−.31	−.07

**Significant at the .01 level.

pothesis that the two correlation coefficients for each behavior rating were obtained from a common population is tenable, the weighted average of these correlations was obtained by averaging the corresponding Z' values. The correlation coefficients based upon the combined data show the following relationships between the S scores and the ratings: negativism, $r = +.35$; assertiveness, $r = +.20$; inadequacy, $r = +.05$; self-distrust, $r = -.26$.

In the group of 59 Ss, S and R minus S were found to be closely correlated ($r = +.67$). The degree of relationship between R minus S and the ratings is also shown in Table 1. The highest correlation was obtained between R minus S and self-distrust ($r = -.31$), indicating that Ss who are judged to be relatively self-confident tend to give many responses. The partial correlation technique was used to determine the correlation between the S score and the ratings with the influence of differences in total responsiveness held constant. The resulting partial correlations showed the following relationships: negativism, $r = +.34$; assertiveness, $r = +.14$; inadequacy, $r = +.18$; self-distrust, $r = -.07$. Since a prediction can be made concerning the probable direction of the correlation, a single-tailed test of significance was used. None of the last three values of r met the criterion of significance at the .05 confidence level. The correlation coefficient of $+.34$ between S and negativism was significant at the .01 confidence level, which indicates a significant but slight relationship between an individual's tendency to give S responses to the Rorschach and a tendency toward negativistic behavior, as rated by his teachers.

Discussion

It has been customary in Rorschach work to utilize the experience-type classification as a differential factor in the interpretation of the S response. Examination of the correlation coefficients between the ratings and the S scores in the two experience types shows no evidence to support the Rorschach hypothesis that the type of behavior to be predicted from the S response is dependent on the experience type in which the S response occurs. The differences between the correlation coefficients of S score and the ratings in the introversive group and the correlations of S score and these ratings in the extratensive group were small and statistically insignificant. This finding has obvious implications for clinical diagnostic work. The practice of interpreting the behavioral significance of the S score in terms of experience type deserves to be discontinued until cogent evidence for its justification becomes available.

The obtained positive relationship between the frequency of S responses and ratings of negativistic behavior offers some support for the Rorschach assumption that the S response represents an oppositional behavior tendency. It is obvious, however, from the moderate size of the obtained relationship that the contribution of the S response, taken singly, to the prediction of negativistic behavior would be relatively small. Perhaps, if combined with other Rorschach factors which provide additional valid variance, consideration of the S response would add to the effectiveness of the Rorschach in predicting negativistic behavior tendencies.

The failure to secure positive findings with respect to the relationships between the S response and ratings of assertiveness, self-distrust, and inadequacy feelings could be conservatively interpreted as being merely a failure to disprove the null hypothesis, given the particular conditions of the present experiment. Such a conservative interpretation rests on the assumption that, in contrast to negativism, attitudes of inadequacy and self-distrust are more difficult to evaluate and consequently were not validly measured by the teachers' ratings. However, it may be pointed out that the teachers rated the Ss in terms of observable behavior, rather than of their own conceptions of the trait being rated, and that the behavioral descriptions were of the type commonly subsumed under these trait classifications. It may be further pointed out that the reliabilities of the teachers' ratings of these traits were as high as the reliability of the ratings of negativism. This, of course, does not assure validity of the ratings, but does at least indicate that there was consistency in the teachers' estimates of the behaviors assumed to be representative of these trait classifications. With these considerations in mind, a stronger interpretation of the findings seems warranted, viz., there is no evidential basis for the practice of inferring such traits from the occurrence of the *S* response, and the practice should be discontinued until such evidence is forthcoming.

Summary

The present investigation was designed to study the validity of the inferences drawn from the Rorschach white space response as reflecting an oppositional tendency.

Eighty-one University of Iowa High School students were tested individually with the Rorschach test. Ratings on 59 of the *S*s, 27 of whom presented introversive and 32 of whom presented extratensive experience types, were made by five teachers with respect to negativism, assertiveness, inadequacy feelings, and self-distrust. These ratings served as the independent criterion measures against which the Rorschach hypotheses were tested.

The results obtained were as follows:

1. No influence of the experience type on the nature of the relationship between the *S* response and the behavior ratings was observed.

2. Without reference to experience type, the number of *S* responses showed a significantly positive relationship of moderate degree to the ratings of negativism.

3. Without reference to experience type, no significant relationship was obtained between the number of *S* responses and ratings of assertiveness, inadequacy feelings, and self-distrust.

From the findings it was concluded that:

1. The usefulness of the Rorschach experience-type classification in the interpretation of the *S* response remains unsubstantiated. It is felt that the utilization of the experience type as a differential factor in the interpretation of the *S* response should be discontinued until evidence for its justification becomes available.

2. The finding of a significant relationship between ratings of negativistic behavior and the *S* response represents a partial corroboration of the Rorschach hypothesis of the *S* response as reflecting an oppositional tendency.

3. Since there is no evidential basis for inferring inadequacy or self-distrust feelings from the *S* response, it is felt that this practice also should be discontinued until such evidence is forthcoming.

Received April 1, 1953.

References

1. Beck, S. J. *Rorschach's test.* Vol I. *Basic processes.* New York: Grune & Stratton, 1944.
2. Benton, A. L. The experimental validation of the Rorschach test. *Brit. J. med. Psychol.*, 1950, **23**, 45-58.
3. Boss, M. Psychologisch-charakterlogische Untersuchungen bei antisozialen Psychopathen mit Hilfe des Rorschachschen Formdeuteversuches. *Z. ges. Neurol. Psychiat.*, 1931, **133**, 544-575.
4. Ebel, R. L. Estimation of reliability of ratings. *Psychometrika*, 1951, **16**, 507-524.
5. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
6. Endacott, J. L. The results of 100 male delinquents on the Rorschach ink blot test. *J. crim. Psychopath.*, 1941, **3**, 47-50.
7. Fonda, C. P. The nature and meaning of the Rorschach white space response. *J. abnorm. soc. Psychol.*, 1951, **46**, 367-377.
8. Hertz, M. R. *Frequency tables for scoring responses to the Rorschach inkblot test.* (3rd ed.) Cleveland: Department of Psychology, Western Reserve Univer., 1951.
9. Hull, C. L. The computation of Pearson's *r* from ranked data. *J. appl. Psychol.*, 1922, **6**, 385-390.
10. Ingram, W. Prediction of aggression from the Rorschach test. Unpublished doctor's dissertation, Northwestern Univer., 1951.
11. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1946.
12. Rabin, A. I. Rorschach test findings in a group of conscientious objectors. *Amer. J. Orthopsychiat.*, 1945, **15**, 514-519.
13. Rapaport, D., Gill, M., & Shaffer, R. *Diagnostic psychological testing.* Vol. II. Chicago: Year Book Publishers, 1946.
14. Rorschach, H. *Psychodiagnostics.* (Trans. by Lemkau and Kronenberg) (2nd Ed.) New York: Grune & Stratton, 1942.
15. Williams, M. The experimental validation of the Rorschach test: I. Experimental correlations. *Amer. J. Orthopsychiat.*, 1952, **22**, 749-754.
16. Zulliger, H. *Jugendliche Diebe im Rorschach Formdeuteversuch. Eine Seelenkundliche und Erzieherische Studie.* Bern: Haupt, 1938.

LIBRARY
1.9.70
J 288

Journal of Consulting Psychology
Vol. 18, No. 1, 1954

Acceptance of Self as Measured by Interviews and the Index of Adjustment and Values[1]

Robert E. Bills
University of Kentucky

Several studies have described the reliability and validity of the Index of Adjustment and Values [1, 4], its ability to differentiate personality types [2], and its usefulness as a research tool [3]. In order to relate research findings on the Index to the larger body of research concerning "acceptance of self" as an important personality variable, it should be established that acceptance of self as measured by the Index and by interviews are essentially the same.

Verbatim transcriptions of recorded, 30-minute, open-ended interviews were collected from 13 people. Five of the subjects, all male, were graduate students in educational administration, and eight, six male and two female, were public school administrators.

The subject's total response between two interviewer responses was accepted as the unit to be scored, and the responses were classified according to content and attitude. Statements were classified according to self, others, work, environment in general, and unclassifiable, and type of attitude (positive, negative, and neutral). Ambivalent statements were scored once for plus and once for minus if neither the positive nor the negative aspects appeared predominant, but if either the positive or negative aspects were stronger, then that aspect was given an additional score of plus or minus.

Two judges, working independently of each other, scored all the interviews. Agreement between judges was determined by checking the scoring of each item and noting the number of agreements and disagreements. The percentages of agreement between the judges were 69, 70, 67, 71, 80, 68, 69, 59, 70, 70, 76, 77, and 63, which are satisfactory for the purposes of this study.

When the judges disagreed on the scoring they discussed the difference and arrived at a decision regarding the unit. The revised scoring was used to determine acceptance of self which was calculated from the formula: number of positive attitudes toward self divided by the total of the positive and negative attitudes toward self.

Subjects were ranked according to acceptance of self shown by the interview material and the Index of Adjustment and Values, and the ranks were correlated to give a rho of .84, which is significantly different from zero at less than the .01 level and which permits the conclusion that what a subject says about himself in an interview corresponds highly with the ratings he gives himself on the Index of Adjustment and Values.

Received April 6, 1953.

References

1. Bills, R. E., Vance, E. L., & McLean, O. S. An Index of Adjustment and Values. *J. consult. Psychol.*, 1951, **15**, 257-261.
2. Bills, R. E. Rorschach characteristics of persons scoring high and low in acceptance of self. *J. consult. Psychol.*, 1953, **17**, 36-38.
3. Bills, R. E. A comparison of scores on the index of adjustment and values with behavior in level-of-aspiration tasks. *J. consult. Psychol.*, 1953, **17**, 206–212.
4. Roberts, G. E. A study of the validity of the index of adjustment and values. *J. consult. Psychol.*, 1952, **16**, 302-304.

---

[1] The data for this study were collected as part of a larger study in the improvement of educational administration sponsored by the W. K. Kellogg Foundation. The University of Kentucky project is directed by Robert L. Hopper, Director of the Bureau of School Services. The writer would like to express his thanks to Henry Oppenheim who has assisted with the project and helped score the interview material.

Journal of Consulting Psychology
Vol. 18, No. 1, 1954

Prediction of Aggression from the Rorschach

Winifred Ingram
Northwestern University[1]

The prevailing need for experimental validation of the Rorschach prompted the investigation reported here. It is an attempt to appraise the validity of the Rorschach as a measure of aggression, with specific reference to the white-space response, by predicting behavior in an experimental situation. The precision of an experiment with the Rorschach may be considerably lowered when the entire test is the basis of the predictions, since many of the scores which enter into the whole interpretation have not been verified. However, it has been shown that no greater precision is secured by using single or composite scores. Rather, success has been achieved more often where the total test was used [3]. It is recognized that an evaluation of isolated scores or interpretations is, in effect, an evaluation of the whole test. Moreover, it is apparent that it is not possible to forecast what scores can be held accountable for prediction errors when other sources of error which can be controlled have been eliminated. The prediction of actual behavior in concrete situations presents many difficulties; among the major problems are the definition of actual behavior and concrete situation [7], the forecasting of expected behavior, the time allotted to the experiment and its analysis, and the number and kind of procedures used [5]. The experimenter must be particularly aware of the usual cultural modes of expression. Presumably, the socialized individual will be likely to utilize these forms of behavior rather than some more highly individualized acts. The above factors were given major attention in the planning of this experiment and the analysis of the results.

Rorschach states that "*S* answers always indicate some sort of *tendency to opposition*" [6, p. 39]. Beck's interpretation in general adheres to this formulation, but he has added some nuances to the scoring and interpretations. Particularly, the delineation of "unitary reactions" and "true reversals of figure and ground" implies a suggestion of varying personality reactions. Depending upon the configuration of the entire test, Beck has applied the adjectives obstinate, skeptical, ambivalent, and compulsive to those persons responding with a certain amount and kind of *S* response [1, pp. 52-57; 2]. These traits have been reported to accompany or follow frustration by the many investigators who have experimented with the frustration-aggression hypothesis, and it seems well established now that aggression is a major reaction to frustration. If *S* is an indicator of aggressive tendencies, individuals with an average or above average *S* score should respond to a situation which is frustrating to them with more aggressive behavior, quantitatively and qualitatively, than individuals who do not produce *S*. To test this hypothesis it was proposed to predict the aggressive behavior expected from two groups differentiated on the basis of the number of *S* responses produced. The standard for the quantity of *S* was adopted from Beck, who states "In most healthy individuals a minimum number of *S* percepts — from two to four — will be found" [3, p. 47], and Molish, who found the mean number of *S* in normal subjects to be between zero and four.[2]

For the purposes of this study aggression is viewed as a dimension with socially acceptable behavior at one extreme and socially unacceptable behavior at the other. From among the adjectives used by Beck and Rorschach to describe the characteristics of the trait projected in *S* and the terms used to specify aggressive behavior by the frustration-aggression theorists, the following were selected as most representative of the trait under consideration: self-assertion, *initiative, persistence, resistance,* and *hostility.* The first three were presumed to represent socially acceptable behavior and the last two were considered to belong in the category of socially unacceptable behavior. In addition to these five aspects of aggression, two terms, cooperation and rapport, were selected as representing the opposite of aggression. Operational definitions, an example of which is given below, were derived from observations of behavior in pilot runs of one of the situations, reports of behavior in frustration experiments [11], and familiarity with interviewee behavior. Each set of operations is preceded by dictionary definitions [9].

---

[1] Now at the Neuropsychiatric Institute, University of Michigan. The author wishes to express gratitude to Dr. T. W. Richards for his valuable suggestions and criticisms in the preparation of this study and to Drs. T. W. Zeigler, R. D. Rabinovitch, and Mrs. Lois Withey who read this manuscript and suggested many helpful revisions.

[2] Personal communication from Dr. Herman Molish, research psychologist, Psychology Laboratory, Michael Reese Hospital, Chicago, Illinois.

Assertion

1. the maintaining or defending, as a cause or claim.

2. positive declaration or averment, affirmation; statement asserted; position advanced.

Operation Interview

1. maintains opinion or attitude about himself in the face of opposing opinions.

a. positively by affirming his position, demonstrating or illustrating his position or ability.

b. negatively by denying the accusation.

2. interrupts interviewer to maintain a position about himself.

3. forceful emphasis of the position without accompanying hostility (i.e., a minimal amount of hostility).

Operation Puzzle

1. ignores instructions even when repeated by examiner.

2. repeats same pattern of response even when it does not achieve the goal.

3. asserts his belief in his ability to solve the puzzle.

Other variables were defined in a similar manner.

Method

The present study consists basically of two parts. The first of these is a comparison of two groups of individuals in two life situations in which aggression might be expressed. The second part of the study is concerned with the correlation between aggression as expressed in each of the life situations and in the Rorschach protocol as a whole.

Sample. From the student records in the office of the Dean of Students of a small state college in Ohio 81 male students were selected for testing. Younger students were given priority if their IQ was average or above and there was nothing in their records to indicate emotional maladjustment. It was hoped, by the selection of Negro students, to decrease problems of rapport which might arise because of racial differences between experimenter and subject. Males were expected to be less inhibited in their display of aggression. Younger students with less experience as experimental subjects could be expected to react more naturally. It was felt that a greater degree of cooperation would be insured by eliminating overtly maladjusted students. The IQ range was limited to average or above to decrease the possibility of sparse records.

Certain cultural traits of American Negroes were expected to modify behavior in the experimental situations. An analysis of the subjects' family background and familiarity with their behavior on campus led to the conclusion that respect for authority, expression of feeling by indirection, and a fairly rigid code of etiquette were characteristic for this group. Status might decrease or increase the amount of aggression overtly displayed. The occupational status of parents indicated that the group on the whole was middle class.

Selection of subjects for the final phases of the experiment was on the basis of the number and type of *S* produced. Each test was given a preliminary examination to score *S,* alone, at the end of each day's testing. For the experimental group only subjects who gave two or more reversals, i.e., responses in which the white space was perceived as the figure and the dark or colored portions of the blot as the ground, were selected.

Approximately half of the 38 subjects tested produced some *S.* Eleven of these met the requirements. These students were then paired on the basis of age and Otis IQ with those who produced no *S.* Three pairs were dropped from the group because of leaving school or erroneous scoring of *S* in their protocols. The final sample consisted of 16 subjects divided equally between the groups. The mean difference in IQ between pairs was three, with a range from nine to five. The mean difference in age was 11 months with a range of 0 to 26 months. The experimental group ranged in age from 18 to 23 (median 19.5) and in Otis IQ from 99 to 117 (median 108.5). The control group ranged in age from 18 to 22 (median 19.5) and in Otis IQ from 99 to 112 (median 108.5). The largest proportion of the total sample was freshmen.

Procedure. The Rorschach test was administered in the usual manner with no exceptions. After all experiments had been completed, the tests were scored by the author according to the Beck method. The scored records and protocols were typed for mailing to the judges.

Two months elapsed between the administration of the Rorschach and the two experimental situations. Two situations were planned which differed in the type of conflict induced; one of them was designed to simulate as closely as possible a life situation. Behavior was recorded by two observers in the first situation (problem solving) and by observers and wire recordings in the second (interview). Unsupervised observers in the first situation were guided in recording behavior by a detailed schedule. In the interview situation the author interviewed while a hidden observer took notes and operated the wire recorder. These two practiced keeping running notes of reactions on two occasions before the subjects were seen.

The problem-solving situation, designed to elicit reaction to intellectual frustration by setting up a goal difficult to achieve, preceded the interview by three to seven days. The Seashore pyramid puzzle [8], selected for this purpose, consists of a three-peg board and six graduated disks with holes which allow them to be slipped easily on the pegs. The goal of the puzzle is to move all six disks from the left-hand peg to the extreme right-hand peg, moving only one disk at a time and never placing a larger disk on a smaller one. Two psychology students served as examiner and observer; the examiner gave the instructions and kept the time and a running record of the subject's behavior, while the observer counted the number of moves and interruptions. A time limit of 30 minutes was set, but the subject was not informed of this limit unless he used all of the time. He was then informed "Your time is up." When the subject failed to observe instructions, the examiner reminded him by repeating pertinent parts of the initial instructions. Questions were answered in the same way. If the subject completed the test in less than 30 minutes, the examiner waited not less than five minutes for the subject to ask to repeat the test. If within this time the subject did not volunteer to repeat, the examiner asked him if he would like to do it again and accepted whatever answer was given. Throughout the situation the examiner exerted a minimum of pressure on the subject.

It was assumed that in the interview situation the interviewer blocked the path to the subject's goal, his vocational choice. Preceding the interview a questionnaire concerning the subject's academic history, vocational choice, extracurricular activities, and family constellation was mailed with a letter to each student stating that the author had completed analysis of the Rorschach tests and wished to interview him. It was requested that the completed questionnaire be returned at the time of the interview [4].

The interview office contained a small closet. The observer sat in this closet, operated the wire recorder, and kept a running record of the subject's behavior. The subject sat next to the interviewer's desk. As the subject was admitted, the observer started the wire recorder so that the sound could become part of the room noises. The interview was usually opened by the interviewer asking for the questionnaire, sometimes by the subject offering it. The interviewer would then read through the questionnaire rapidly and ask the subject, "What difficulty are you having with your classes?" Next the interviewer requested "Tell me more about why you would 'like to be a ....,' " naming the subject's vocational choice as stated in the questionnaire. Following the subject's response, the interviewer pretended to read the subject's Rorschach report and then stated seriously: "After considering your records carefully, your Rorschach, the report of your puzzle performance, and what you have said here, it's difficult to see how you're making the grade. It looks like a pure waste of time to me." A concealed stop watch was started as this statement was completed, and the time of the subject's first response, verbal or otherwise, was recorded. When the interviewer felt certain that the response to this statement was completed, the time of the last response was recorded. The interview was terminated with an explanation of its purpose; questions were answered and a thorough attempt was made to desensitize the subject before he left.

The three situations provided settings for different roles for the participants, observers, and subjects. Use of the Dean of Student's office and its services gave the author quasi-official status, which was further enhanced by a notice placed on the bulletin board that the author, an employee of the U. S. Public Health Service, was engaged in making a survey of normal young men. The role of an objective psychologist was developed by not using the subjects' names, with the explanation that this enabled the author to remain more objective. Living on campus facilitated friendly relationships. It was assumed that establishing such rapport would reinforce the frustrating effect of the belittling statement introduced in the interview. Although the assistants in the puzzle situation were cast as disinterested observers, there was evidence that a few of the subjects felt some personal involvement with the observers. This is not surprising since the student body is small. Nevertheless it is felt that the subject was pitted against himself rather than against the others in the situation.

Each category of aggression and the assumed opposites were set up as an independent five-point rating scale; all seven scales were printed on a single page. Rorschach, puzzle, and interview protocols were rated by means of the same scales. Judges of the Rorschach records were asked to arrive at an interpretation of the whole record and to record their impression of each subject on the rating scales. There was no training period; instructions and records were mailed. Judges of the Rorschach were selected primarily on the basis of their training in the Beck system.[3] Judges of the other two situations were selected because of their familiarity with rating procedures,[4] a separate group rating each situation. The author acted as a judge for all situations. Definitions and instructions were reviewed for each judge before the ratings were performed. The protocols were identified only by case number, and these were masked before the author acted as a judge.

Results

Results of the statistical analysis of data will be presented in the following order: (*a*) reliability of ratings of aggressive behavior in the life situations; (*b*) significance of differences between groups high and low in S; (*c*) ratings of aggression expressed in the Rorschach performance; (*d*) correlations between ratings of aggression in the life situations and ratings of aggression on the Rorschach records.

Reliability of ratings of aggressive behavior in the life situations. Reliability coefficients were calculated using the total group of 18 subjects.[5] Each coefficient represents intercorrelation between three judges for the puzzle situation and a different group of three judges for the interview situation. For persistence and hostility in the puzzle situation and for initiative and hostility in the interview situation p equaled .01. For initiative, cooperation, and rapport in the puzzle situation and for assertion, persistence, and rapport in the interview situation p was .05. These findings indicate that of the seven behavior variables under con-

Table 1

Reliability of Ratings

($N = 16$)

Behavior category	Situation Puzzle	 Interview	 Rorschach
Assertion	.15	.58*	.74**
Initiative	.55*	.64**	.70**
Persistence	.59**	.51*	.71**
Resistance	.11	.41	.65**
Hostility	.73**	.71**	.57*
Cooperation	.49*	.27	.72**
Rapport	.56*	.47*	.44

* Significant at .05 level.
** Significant at .01 level.

---

[3] Drs. Seymour Klebanoff, Herman Molish, Alan Rosenwald, and Meyer Williams acted as Rorschach judges.

[4] Miss Alene Gustavson and Dr. Harry Klehr rated interview records. Dr. Ross Morgan and Mr. Dave Murray rated puzzle records.

[5] Calculated by a method suggested by Dr. E. L. Clark.

sideration, four (initiative, persistence, hostility, and rapport) were measured reliably by both the puzzle and interview situations; cooperation was measured reliably only in the puzzle situation and assertion only in the interview situation.

Significance of differences between groups high and low on S. In order to test the significance of differences between groups, ratings for each subject were averaged and the usual t test was made. A significant level of differentiation ($p = .05$) between groups was obtained in the interview situation on the following variables: initiative, resistance, hostility, cooperation, and rapport. Differentiation at the same level was obtained in the puzzle situation on assertion and hostility.

Limiting discussion to differences at the .05 level of confidence or better: (*a*) in the interview situation the group high in *S* response on the Rorschach exhibits more initiative, more hostility, but also more cooperation and rapport and less resistance; (*b*) in the puzzle situation the group high in *S* response exhibits less assertion and less hostility.

So far as the interview situation is concerned, the original hypothesis is confirmed, namely that individuals high on *S* response are more aggressive. They are aggressive in being generally more participatory and interactive. However, in the puzzle situation (primarily a situation of intellectual challenge involving social interaction only peripherally) the high *S* individuals are not aggressive. There is a strong suggestion that the two situations differ markedly in the degree to which they elicit aggressive behavior as measured in this study.

Ratings of aggression expressed in the Rorschach performance. Intercorrelations between judges rating the total Rorschach record are shown in Table 1. A probability level of .01 was obtained for the intercorrelations on the variables assertion, initiative, persistence, resistance, and cooperation. The intercorrelations for hostility were significant at the .05 level. These coefficients suggest that the raters were in considerably greater agreement in rating the Rorschach than they were in rating either the interview or the puzzle situations.

Correlation between ratings of aggression in the life situations and ratings of aggression on the Rorschach record. Coefficients of correlation between the ratings of aggression on the total Rorschach and those for each of the life situations are shown in Table 2. The correlation for the variable, persistence, in the puzzle situation is significant at the .05 level of probability. The probability approached the .05 level for the variable, resistance, in the interview situation. A general conclusion is that there is little relationship between the aggression inferred in either situation and that inferred from the total Rorschach. The largest coefficient (.52) suggests a slight correspondence between the puzzle and Rorschach situations for persistence, in the direction which is perhaps expected. In the interview situation, however, the largest coefficient (—.40) suggests that those who seem most resistant on the Rorschach seem least resistant in the interview situation.

Table 2

Correlation between Ratings of the Life Situations and the Rorschach

	Situation	
Category	Puzzle	Interview
Assertion	.34	—.37
Initiative	—.11	.21
Persistence	.52*	.35
Resistance	—.17	—.40
Hostility	—.08	.27
Cooperation	.19	—.01
Rapport	—.30	—.11

* Significant at .05 level.

Summary and Conclusions

In order to demonstrate that the interpretation of the Rorschach white space score, *S*, as aggression is valid, 16 Negro males between the ages of 18 and 24 were subjected to intellectual and interpersonal frustration after having been given the Rorschach test. The resultant observations of behavior in two situations, a problem-solving test and an interview, and the scored Rorschach tests were subjected to a rating procedure on a five-point scale of aggression, composed of five behavior variables, and a nonaggressive scale, composed of two behavior variables.

Of the five aggressive variables the most reliably rated were initiative, persistence, and hostility. Of the two nonaggressive variables rapport was the more reliably rated.

The major conclusions from the study are:

1. The high *S* group was more aggressive (in a participatory, interactive way) in a frustrating interview situation than the low *S* group.

2. The high *S* group was no more aggressive than the low *S* group in a situation of intellectual challenge in which social interaction was at a minimum.

3. The judges' impressions of aggression as inferred from the total Rorschach record were but slightly related to aggression as measured in the interview and puzzle situations.

Received May 18, 1953.

References

1. Beck, S. J. *Rorschach's test.* Vol. I. *Basic processes.* New York: Grune & Stratton, 1944.
2. Beck, S. J. *Rorschach's test.* Vol. II. *A variety of personality pictures.* New York: Grune & Stratton, 1946.
3. Cronbach, L. J. Statistical methods applied to Rorschach scores. *Psychol. Bull.,* 1949, 46, 393-429.
4. Murray, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.
5. OSS Assessment Staff. *Assessment of men. Selection of personnel for the Office of Strategic Services.* New York: Rinehart, 1948.
6. Rorschach, H. *Psychodiagnostics.* New York: Grune & Stratton, 1942.
7. Sargent, Helen. Projective methods; their origin, theory and current applications in personality research. *Psychol. Bull.,* 1945, 42, 257-293.
8. Seashore, R. S. Pyramid puzzle—a useful device for studying thought. *Amer. J. Psychol.,* 1938, 51, 549-557.
9. *Webster's New International Dictionary of the English Language.* (2nd Ed., Unabridged) Springfield, Mass.: G. & C. Merriam, 1934.
10. Yacorzynski, G. K. An experimental study of the functions of the frontal lobes in man. *Psychosom. Med.,* 1945, 7, 97-107.
11. Zander, A. F. A study of experimental frustration. *Psychol. Monogr.,* 1944, 56, No. 3 (Whole No. 256).

Journal of Consulting Psychology
Vol. 18, No. 1, 1954

The Effects of Figure-Ground Contrast Upon Perception as Evaluated by a Modified Rorschach Technique[1]

Arvord W. Belden and E. Earl Baughman
University of Wisconsin

In a recent article [1] results of research designed to clarify the role of stimulus attributes in determining perceptual behavior on the Rorschach test were presented by the second author. Among other things, this research demonstrated a significant increase in perceived movement responses (scored as *M* responses according to Rorschach convention) on a silhouette series of blots in which figure-ground contrast had been maximized. The exact relationship between figure-ground contrast and movement responses, as well as relationships with other dimensions of perception exhibited on the Rorschach test, could not be specified, however. This was due to the fact that there was no systematic variation of contrast in the experiment, but rather only a utilization of stimuli in which contrast was either eliminated or maximized. Examples of these polar extremes can be seen in Fig. 1, where they are indicated as Cards A and G.

The purpose of the present paper is twofold: (*a*) to present data which make possible a systematic analysis of the relationship between figure-ground contrast and perception in a Rorschach type situation, and (*b*) to present an example of how methodology employed with projective techniques may be combined fruitfully with more traditional experimental methods toward the goal of a more complete understanding of perception.

Method

Stimulus materials. A graded series of seven stimulus figures was used in the experiment, the figures differing from one another only in terms of the amount of light reflected. The form of each figure was identical to that of Card II in the Rorschach test. Card II was selected as the basic form for this study because of its demonstrated ability to evoke movement

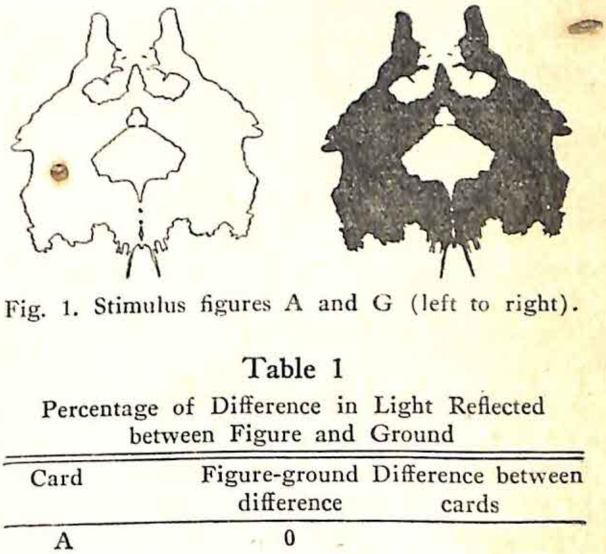


Fig. 1. Stimulus figures A and G (left to right).

Table 1

Percentage of Difference in Light Reflected between Figure and Ground

Card	Figure-ground difference	Difference between cards
A	0	
		13.1
B	13.1	
		18.6
C	31.7	
		22.0
D	53.7	
		22.5
E	76.2	
		15.0
F	91.2	
		5.7
G	96.9	

---

[1] The authors wish to thank Drs D. W. Fiske, D. A. Grant, G. Calden, and H. A. Page for their criticisms pertaining to this article. Appreciation is also expressed to Drs. F. A. Mote and H. Leibowitz for assistance in constructing materials used in the experiment.

responses [1]. The two extremes of the seven stimulus cards are presented in reduced size in Fig. 1. Table 1 indicates the percentage of difference in light value between figure and ground for each card as measured by a Macbeth illuminometer.

Cards B through G were prepared by covering the figure on Card II in the standard Rorschach with a solid coat of black India ink (in order to eliminate color, shading, and internal form effects) and then photographing the card. Actual size prints were then made with different contrast values. As can be seen in Table 1, the difference in contrast between Cards F and G was less than that between Cards B-C, C-D, D-E, and E-F, the latter cards differing from one another in approximately equal step intervals. Card G was included, however, in order to provide a card with maximum figure-ground contrast within the limitations of the photographic process used. Card G will be referred to hereafter as the figure with maximum contrast, and Card B will be referred to as the figure with minimum contrast. Cards F through B were identical to Card G in Fig. 1 except for systematic decrements in figure-ground contrast.

Card A is termed a peripheral line figure and was included in the series in order to provide a stimulus without figure-ground contrast, the white of the figure being the same as that of the ground. It was created by placing a transparent paper over the original Rorschach card and tracing in such a way that the pen was always in contact with both the white ground and either a shaded or colored portion of the card. Photographing then followed the procedure outlined above.

Experimental design and subjects. A total of 140 subjects was used in the experiment, 20 subjects being randomly assigned to each of the seven cards. All subjects were white male students enrolled in an introductory psychology course at the University of Wisconsin. After the composition of the groups had been determined, the groups were compared in terms of age, intelligence, and educational level since these variables are known to affect Rorschach performance. Analysis of variance revealed no significant differences between the groups with respect to these variables.

Testing procedure and treatment of data. All subjects were tested by the first author in order to eliminate examiner differences [2]. The entire testing situation was recorded by wire recorder to insure complete and accurate protocols. Uniform instructions, adapted from Beck [3], were given to each subject as follows:

> You will be given a card with a design on it. Look at the card and tell me what you see or anything that might be represented there. Look at the card as long as you like, only be sure to tell me everything that you see on the card as you look at it. When you are finished with the card lay it down on the desk as a sign that you are through with it.

No encouragement for greater productivity was given after the first response. Upon completion of the free association a standard inquiry was conducted. Each protocol was then summarized by the first author following Beck [3], except for the addition of Klopfer's *FM* score [7]. Seventy-six responses were also scored independently by the second author, perfect scoring agreement being found on 88 per cent of the responses. The scoring categories utilized are listed in Table 2.

The data were analyzed in terms of group comparisons, each group being compared with every other group in respect to each of the scoring categories. Fisher and Yates's [5] exact method was used to test the significance of differences in frequency between groups. The theoretical frequencies used in this procedure were determined by casting the scores of all 140 students into total distributions for each of the scoring categories and choosing a cutting point as close to the median value as possible. It was necessary to use this statistical technique because of two factors: (*a*) the marked skewness of the distributions of most Rorschach scoring categories and (*b*) the small theoretical frequencies in some categories. Such difficulties in the statistical handling of Rorschach data have been reviewed by Cronbach [4].

Two exceptions in the above handling of the data are to be noted, however, involving: (*a*) average time per response and (*b*) average total time per subject. The characteristics of these distributions were such as to permit an analysis of variance technique. This analysis was carried out, in addition to the exact method, for the above two scoring categories.

Results

A table presenting median and mean scores for each group in each scoring category may be secured from the American Documentation Institute.[2] Table 2 indicates significant differences between groups for each scoring category when each group was compared with every other group by the exact method. Figures 2, 3, 4, and 5 present graphs of the relationship between degrees of contrast and 11 of the 16 measures of perceptual behavior used in this study. Five measures (*D*, *Dd*, *P*, *FM*, and *M* + *FM*) are not graphed since they revealed linear relationships with approximately zero slope comparable to *W* and *M* as seen in Fig. 2. Card A is left discontinuous with the remaining cards in Fig. 2, 3, 4, and 5 for two reasons: (*a*) to emphasize the marked difference between this card and the remaining cards as perceived by the subjects, and (*b*) to make more clear the orderly relationship between degree of contrast and perceptual behavior as long as at least a minimum brightness contrast is maintained. Figures 4 and 5 were constructed on the basis of mean scores for the time categories since the characteristics of these distributions were such as to suggest that the means would be more reliable measures than the medians.

Table 2

Significant Differences between Groups According to Scoring Categories

Scoring category*	Significance level .01	Significance level .05
R		A–B
D		A–B, A–G
M	B–A, D–A, F–A	C–A
FM		E–C
Total (–) Form		A–B, A–C
P	B–A, C–A, D–A, F–A	E–A, G–A
S	A–B	A–D, A–F
T/1R		F–C, F–G
T/R	B–E	B–A, B–D
T. Time		B–C

Note.—The first letter in each entry indicates which of the two groups exceeded the other with respect to the scoring category. For example, A exceeded B in productivity (R).

*No comparisons for W, Dd, M + FM, FY, Total (+) Form, and Content Categories reached the .05 level of significance; hence these scoring categories are not listed in the table.

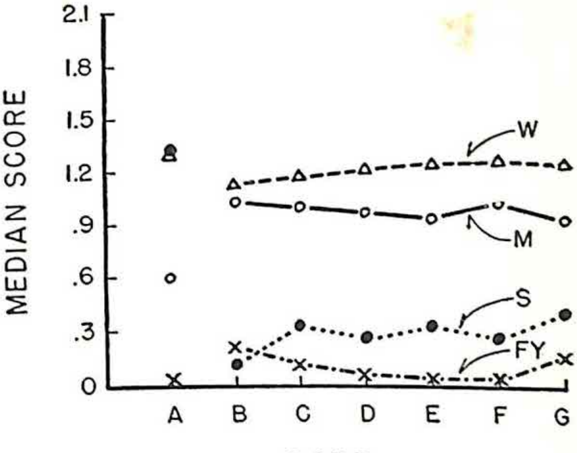


Fig. 2. Median scores for *W*, *M*, *S*, and *FY*.

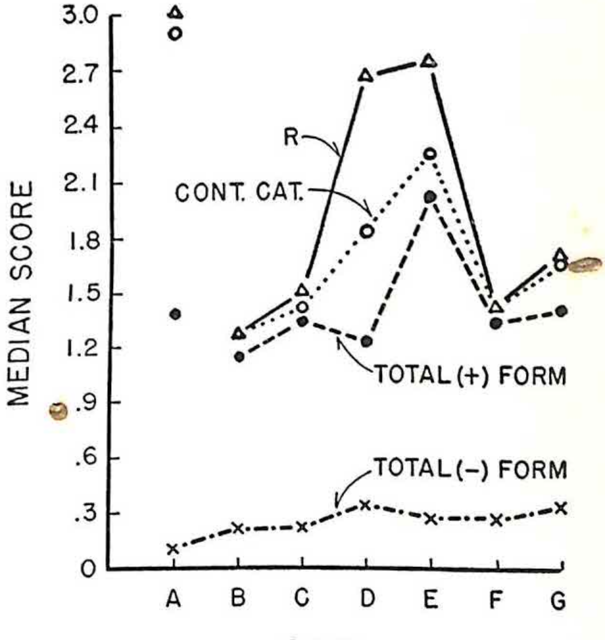


Fig. 3. Median scores for *R*, Content Categories, Total (+) Form, and Total (—) Form.

Nineteen of the 25 entries in Table 2 involve Card A, thus emphasizing the marked changes occurring in perceptual behavior when brightness contrast between figure and ground is eliminated. This is best seen by the fact that the most frequent response given to this card, two humans in some type of activity, dimin-

---

[2] Order Document No. 4076 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photoprints. Make checks payable to Chief, Photoduplication Service, Library of Congress.

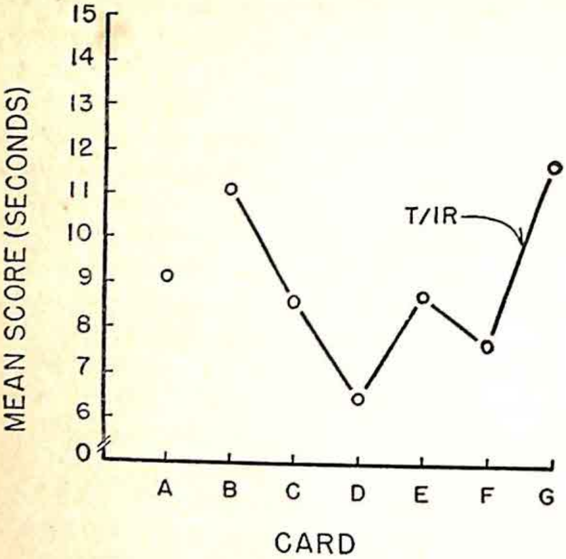


Fig. 4. Mean score for $T/1R$.

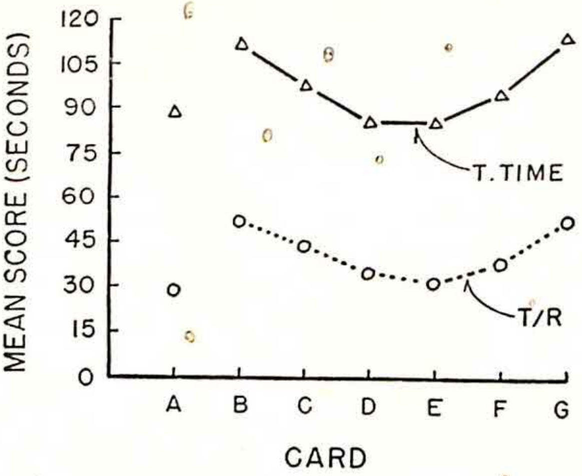


Fig. 5. Mean scores for T/R and Total Time.

ishes markedly in frequency on Card A (see *P* and *M* scores in Table 2). In addition, the subject is much more likely to restructure figure-ground relationships on Card A (see *S* score in Table 2) than on the remaining cards. These findings are consistent with previous results obtained when figures comparable to Cards A and G were utilized [1].

Our primary interest from the beginning was with the relationships which would unfold on Cards B through G. On these cards the figure is defined for the subject by means of a brightness contrast, the only difference between the cards being in terms of degree of brightness contrast. Table 2 and Fig. 2 and 3 point to no significant or systematic relationships between contrast and the following scoring categories: *W, D, Dd, M, FM, M + FM, P, S,* and Total (−) Form.[3] Only *W, M,* and Total (−) Form are graphed to illustrate these relationships, thus showing that they may be described essentially by a straight line with zero slope. Figure 3 indicates, however, an increase in three scores (*R,* Content Categories, and Total [+] Form) as contrast increases, until a maximum is reached on Card E, following which there is a decrease in these scores with further increase in contrast. Figures 2, 4, and 5 indicate a reverse type relationship between contrast and four other scores: Total Time, $T/1R$, T/R, and *FY*. These scores reach a minimal point on Card E (except for $T/1R$ which reaches a minimum on Card D) and then increase with further increase in contrast.

When T/R scores were subjected to an analysis of variance, the resulting F was found to be very significant ($F = 5.20$, where $F = 2.95$ is significant at the .01 level). An $F = 2.06$ was obtained when Total Time scores were subjected to a similar analysis, where an $F = 2.17$ would have been significant at the .05 level.

Discussion

Since previous research [1] had demonstrated a marked increase (in comparison with standard Rorschach forms) in movement responses (*M*) on a series of blots with contrast values comparable to those of Card G in the present study, it was anticipated that an increase in *M* responses would be demonstrated corresponding to an increase in contrast values. This relationship was not found, however. A possible explanation for this resides in the uniformally consistent level at which movement responses were evoked by the particular form which was chosen for this study. Thus, the range of behavior with respect to movement responses proved to be too constricted for an adequate test of this hypothesis. The need for further work with diverse forms providing wider variations in perceptual behavior is clearly indicated.

---

[3] All scores follow accepted Rorschach convention [3] except for Total (+) Form, Total (−) Form, and Content Categories. To obtain Total (+) Form and Total (−) Form scores, all responses were evaluated as to accuracy or inaccuracy of form perceived even though secondary determinants may have been utilized. This was necessitated by the small sample of behavior secured from each individual. Content Categories refers to the number of different types of content perceived by the subject.

The systematic relationships pictured in Fig. 2, 3, 4, and 5 between contrast and seven of the scores used to evaluate perceptual behavior form the basis of an interesting hypothesis. These curves reveal that more responses (R) are given in shorter time (T/R, $T/1R$, Total Time) with greater accuracy (Total [+] Form) and with greater variety in content (Content Categories) when figure-ground contrast is intermediate in value rather than minimal or maximal as defined in this study. Moreover, if one accepts the validity of the usual Rorschach interpretation of FY responses, the data suggest that feelings of uneasiness or anxiety may diminish at intermediate contrast values. The data are so limited in the FY category, however, that only a very tentative suggestion in this direction is justified.

Consideration of the above relationships immediately suggests the idea that perceptual behavior is most efficient when figure-ground contrast is at a level comparable to that provided by Card E (a level at which the figure reflects approximately 25 per cent as much light as is reflected by the ground). As one either increases contrast beyond this point, or decreases it, perceptual behavior tends to become less efficient. This interpretation of the data is consistent with a definition of efficient behavior which stresses maximal output per unit time at a high level of accuracy concomitant with pleasant feeling tone.

The methodology used in this study differs from that of the more usual experimental approach in which both the stimulus conditions and the subject's responses are rather rigidly controlled. In this experiment we have chosen instead to control carefully the stimulus conditions but to make multidimensional measurements of free or unstructured responses by the subjects. This is equivalent to saying to the subject, "If this stimulus condition affects you then you are free to show us how it affects you," rather than, "If this stimulus affects you then it must affect you through the behavior which we have already selected."

An obvious criticism arises immediately, however, in the sense that the many measures used to evaluate the behavior are not measures of independent functions. This problem has been discussed elsewhere by Fiske and Baughman [6]. By ferreting out these interrelations before variations are introduced into the stimulus conditions we should be able to overcome these difficulties. Thus, a previous study [6] has shown that as productivity (R) increases there is a correspondingly regular increase in $F+$ and $F-$ responses. Consequently, the fact that we get such an increase in $F+$ responses in the present study but *not* a similar increase in $F-$ responses makes more meaningful the interpretation which we have given to the data.

It is possible, of course, that our results are due to the particular form which we chose to work with or to unidentified differences between our groups. Further research will be necessary to determine whether the concept of perceptual efficiency, in the sense that we have proposed it, will hold with other groups of subjects and stimulus forms.

Summary

Seven figures were constructed comparable to Card II in the Rorschach test with respect to form, but differing from one another in degree of figure-ground contrast. Each of these seven cards was administered to a different group of 20 male subjects who were randomly assigned from a total of 140 subjects in elementary psychology courses. Administration and scoring of protocols followed accepted Rorschach procedure. Each group was then compared with every other group in order to determine the relationship between contrast and perceptual behavior, with the following findings:

1. Nineteen out of 25 significant differences in Rorschach responses between cards involved Card A on which there was no brightness contrast between figure and ground.

2. When Cards B through G were considered (on which approximately equal step increases in contrast were present) no relationship between contrast and the following factors was found: W, D, Dd, M, FM, $M+FM$, Total ($-$) Form, P, and S.

3. When Cards B through G were considered, R, Content Categories, and Total ($+$) Form were found to increase with increasing

contrast until a maximum was reached on Card E; these scores then decreased with further increase in contrast.

4. When Cards B through G were considered, *T/1R, T/R,* Total Time, and *FY* were found to decrease with increasing contrast until a minimum was reached (usually on Card E); these scores then increased with further increases in contrast.

These findings were interpreted in terms of the concept of perceptual efficiency. It was suggested that perception is most efficient at a point of contrast which falls between the minimum and maximum as defined in this study, a point at which the figure reflects approximately 25% of the light which is reflected by the ground.

Received March 30, 1953.

References

1. Baughman, E. E. A comparative analysis of Rorschach forms with altered stimulus characteristics. *J. proj. Tech.,* in press.
2. Baughman, E. E. Rorschach scores as a function of examiner differences. *J. proj. Tech.,* 1951, 15, 243-249.
3. Beck, S. J. *Rorschach's test.* Vol. I. *Basic processes.* New York: Grune & Stratton, 1949.
4. Cronbach, L. J. Statistical methods applied to Rorschach scores: a review. *Psychol. Bull.,* 1949, 46, 393-429.
5. Finney, D. J. The Fisher-Yates test of significance in 2 × 2 contingency tables. *Biometrika,* 1948, 35, 145-156.
6. Fiske, D. W., & Baughman, E. E. Relationships between Rorschach scoring categories and the total number of responses. *J. abnorm. soc. Psychol.,* 1953, 48, 25-32.
7. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1946.

Journal of Consulting Psychology
Vol. 18, No. 1, 1954

Interrelationships Among Several Measures of Anxiety and Hostility[1]

Leonard D. Goodstein
State University of Iowa

The problem of anxiety and its measurement has become increasingly important in contemporary psychology in both clinical and experimental work.

The Taylor Manifest Anxiety Scale. The Taylor Manifest Anxiety Scale (*A* scale) [8, 9, 24, 25] has been widely employed as a measure of anxiety. The *A* scale, initially, consisted of 65 MMPI items that were rated by judges to be indicative of anxiety; those items found to be nondiscriminating, through item analysis, were subsequently eliminated. The present scale consists of 50 such items.

Validity studies of the *A* scale have been of three types. A number of experimental studies have assumed that *A*-scale scores are an index of the *S*s' state of *reactivity* or *excitability,* which in turn reflects their general level of drive (*D*). Using Hull's [12] theoretical formulation relating response strength to drive, these studies have predicted performance in a variety of laboratory experiments, involving conditioning [1, 10, 16, 21, 22, 24], reaction time [18, 27], verbal learning [15, 17, 23, 26], as well as several more complex behavioral situations [5, 19, 20, 28, 29].

Secondly, several attempts have been made to relate *A*-scale scores to manifestations of clinically diagnosed anxiety states. Gleser and Ulett [6] reported a correlation of .61 between *A*-scale scores and psychiatrists' ratings of anxiety proneness. Lauterbach [14] found a correlation of .31 (significant beyond the 1% level of confidence) between *A*-scale scores and psychologists' ratings of overt anxiety, but the correlation with psychiatrists' ratings was not significantly different from zero.

Thirdly, additional information on the clinical validity of the *A* scale is found in those studies reporting the relationships between *A*-scale scores and other tests purporting to measure anxiety and other personality variables. Lauterbach [14] reported a correlation of .62 between the *A* scale and Welsh's Anxiety Index, an empirically derived MMPI anxiety measure, while Holtzman *et al.* [11] have reported a correlation of .86 with the Winne Scale of Neuroticism, another empirically derived MMPI measure.

The present study is of the third type. It represents an attempt to study the relationship between *A*-scale scores and two other measuring instruments, the Elizur Rorschach Content Test and a new, multiple-choice form of the TAT.

The Elizur Rorschach Content Test. A relatively new technique for scoring Rorschach protocols, based on an analysis of protocol content, has been developed by Elizur [3]. Rorschach responses containing clear-cut evidence of strong anxiety, such as a frightening giant, are given score credit of 2, while responses containing less clear-cut evidence, such as an unpleasant animal, are given score credit of 1. A quantitative score expressing the amount of anxiety is then obtained by totaling the score credits for a particular protocol. A hostility score is obtained in a similar fashion.

Elizur [3] found significantly high correlations between his Rorschach anxiety scores and indicators of anxiety derived from interviews, questionnaires, and rating scales obtained from normal subjects. His anxiety score was also found to differentiate (at the 1% level of confidence) between a group of psychiatrically diagnosed neurotics and a group of normals who had been matched on the basis of other relevant variables.

---

[1] Based on a paper presented at the annual meetings of the Midwestern Psychological Association, Chicago, Illinois, May 1, 1953.

The Iowa Multiple-Choice Picture Interpretation Test. Because of the limitations involved in the usual administration and scoring of the TAT, a simpler method of presentation and a more objective method of scoring was devised as part of a research being carried on at Iowa.

Ten TAT cards, 1, 2, 4, 6BM, 7BM, 7GF, 8BM, 13, 14, and 17BM, were selected on the basis of Eron's [4] normative data and personal experience. Four alternative responses were prepared for each of these pictures; each response was composed of a single statement involving one of the following types of themes: anxiety, hostility, achievement, and blandness.

Card 8GF, a young woman sitting with her chin in her hand looking off into space, was used for the purpose of illustration. The four alternatives to this sample picture were:

A. She feels that she cannot go to the party because no one ever asks her to dance. (*Anxiety*)

B. She is annoyed because she has to work while her friends go out. (*Hostility*)

C. She is looking forward to her opening night as the star of a great show. (*Achievement*)

D. She is listening to her favorite radio program. (*Blandness*)

The *S*s are asked to rank order these alternatives according to the interpretation *they* would be most likely to give. Each area or theme is scored as the sum of the assigned ranks. It should be noted, therefore, that the lower the numerical score, the higher the over-all preference expressed for the particular alternative.

While little is known at the present time about the usefulness of those scores, Daily and Hurley [2], in a preliminary study, demonstrated that subjects scoring high on the achievement variable gave significantly more anticipations in a verbal learning situation than those scoring low in achievement. Hurley [13] has essentially confirmed these results.

Hypotheses

The specific hypotheses were:

1. The three measures of anxiety will be positively related.

2. The two measures of hostility will be positively related.

3. The anxiety measures will be negatively related to the hostility measures, as individuals able to express hostility should be less anxious than those unable to do so.

Subjects and Procedures

The three tests, the *A* scale, the group Rorschach (administered according to Harrower's [7] instructions and scored according to Elizur's [3] system), and the Iowa Multiple-Choice Picture Interpretation Test, were administered to 57 college students enrolled in an undergraduate course at the State University of Iowa.

Results

A complete intercorrelation matrix of test scores (including the total number of Ror-

Table 1

Intercorrelation Matrix among All Test Scores

($N = 57$)

Test	Rorschach total *R*	Elizur Anx.	Elizur Host.	Picture Interpretation Test Anx.	Picture Interpretation Test Host.	Picture Interpretation Test Ach.	Picture Interpretation Test Bland
Taylor *A* scale	.18	.38**	—.20	—.17	—.24	.22	.19
Rorschach total R		.25*	—.25*	.14	.03	—.12	—.02
Elizur anxiety			.07	—.06	—.06	—.02	.14
Elizur hostility				—.08	—.02	.02	.02
Picture Interpretation Test†							
Anxiety					*.11*	*—.60*	*—.57*
Hostility						*—.72*	*—.72*
Achievement							*.46*
Blandness							

* Significant at .05 level of confidence.

** Significant at .01 level of confidence.

† The intercorrelations among the Picture Interpretation Test scores are confounded by the rank-order scoring procedures.

schach responses in each protocol) was constructed and is presented in Table 1.

As may be observed in Table 1, a low but significant positive correlation was found between *A*-scale scores and Elizur anxiety scores, indicating some common variance. A low but significant positive correlation was found between total number of Rorschach responses and Elizur anxiety scores while a negative correlation of the same order was found between total *R* and Elizur hostility scores. This would seem to indicate that *S*s who give many responses tend to give responses scored in the anxiety category while those giving few responses tend to give responses scored in the hostility category.

No significant relationships were demonstrated among the *A*-scale scores and the Rorschach scores on the one hand and the Picture Interpretation Test scores on the other hand. The anxiety and hostility scores provided by the Picture Interpretation Test do not seem to tap the same sources of variance as the other measures.

It is difficult to interpret the intercorrelations among the Picture Interpretation Test scores, as these relationships, at least in part, are due to an artifact of the rank-order scoring procedure.

As the *A* scale has usually been found to be nondiscriminating over the middle range of scores, two groups of *S*s (anxious and nonanxious) were selected from the extremes of the *A*-scale distribution. Each group consisted of 10 *S*s whose *A*-scale scores fell, respectively, within the upper and lower 20% of scores for a standardization population of about 2,500 *S*s. The performance of these two groups was then compared on each of the other measures.

The means and standard deviations for the anxious and the nonanxious groups, the resulting *t*'s, and the probability estimates for the *t*'s are presented in Table 2. In the first four cases single-tailed hypotheses were used since the differences were in the predicted direction. The anxious group tended to give more total Rorschach responses, higher Elizur anxiety scores, lower Elizur hostility scores, and give greater preference to the anxiety foil of the Iowa Picture Interpretation Test than the nonanxious group.

These results are in substantial agreement with those reported in Table 2; however, the results based on the extreme groups are statistically more reliable. The relationship between *A*-scale scores and total *R* is essentially a confirmation of the findings of Westrope [29], who reported that anxious *S*s tended to give more Rorschach responses than nonanxious *S*s in individual Rorschach administrations.

Discussion

The present study is part of a continuing series of investigations on manifest anxiety and its behavioral correlates. To date, the Taylor *A* scale has been the principal instrument utilized in these studies. Recently, however, we have become interested in attempting to tap this source of variance by some other technique. The rather restricted range of *A*-scale scores in a college population, the lack of dis-

Table 2

Means on Seven Measures for Anxious and Nonanxious *S*s

($N = 10$ in each group)

Test	Mean Anx.	Mean N–Anx.	*SD* Anx.	*SD* N–Anx.	*t*	*p*
Rorschach						
Total R	27.4	21.6	7.7	5.0	1.9	<.03*
Elizur anxiety	5.2	2.2	2.3	1.3	3.4	<.001*
Elizur hostility	1.2	3.1	1.3	2.9	1.8	<.04*
Picture Interpretation Test						
Anxiety	23.8	26.4	3.5	4.3	1.4	<.10*
Hostility	27.9	31.7	3.8	5.1	1.8	.10
Achievement	22.2	19.3	4.1	3.4	0.9	.40
Blandness	26.1	23.4	4.3	4.6	1.3	.20

* Using single-tailed hypothesis.

crimination over the middle range of *A*-scale scores, and the susceptibility of the *A* scale to faking [9] have, in part, motivated this interest. It was hoped that the Rorschach or a modified form of the TAT would be helpful in this regard. The present study would seem to indicate that the Rorschach might be somewhat more helpful than this modification of the TAT.

One might question whether the relationships among *A*-scale scores and Elizur scores are merely dependent upon the total number of Rorschach responses. Partial correlations, holding total *R* constant, were computed to answer this question. The partial correlation between *A*-scale scores and Elizur anxiety scores, holding total *R* constant, was .36, while the correlation with the Elizur hostility scores, holding total *R* constant, was —.26. These partial correlations are not significantly different from the whole correlations presented in Table 1. The relationship between *A*-scale scores and the Elizur scores does not seem to be merely dependent upon the total number of Rorschach responses, that is, some general productivity factor.

Summary and Conclusions

1. Not all tests thought to measure anxiety measure the same variable. Significant relationships were found between the Taylor Manifest Anxiety Scale and the Elizur Rorschach Content Test of Anxiety. An anxiety score derived from the Iowa Multiple-Choice Picture Interpretation Test, a modified form of the TAT, was not significantly related to the two other measures of anxiety.

2. Not all tests thought to measure hostility measure the same variable. No significant relationship was found between the Elizur Rorschach Content Test of Hostility and the hostility score derived from the Iowa Multiple-Choice Picture Interpretation Test.

3. The relationship between anxiety and hostility depends upon the measures used. A negative relationship was found between the Taylor Manifest Anxiety Scale and the Elizur Rorschach Content Test of Hostility while a positive relationship was found between the *A* scale and the hostility scores derived from the Iowa Multiple-Choice Picture Interpretation Test. The tests of statistical significance, however, do not permit much confidence in these results.

Received May 11, 1953

References

1. Bitterman, M. E., & Holtzman, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615-623.
2. Daily, J. M., & Hurley, J. R. Need Achievement and behavior in a rote learning situation. Unpublished manuscript, State Univer. of Iowa, 1952.
3. Elizur, A. Content analysis of the Rorschach with regard to anxiety and hostility. *Rorschach Res. Exch.*, 1949, **13**, 247-283.
4. Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64**, No. 9 (Whole No. 315).
5. Farber, I. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120-125.
6. Gleser, Goldine, & Ulett, G. The Saslow Screening Test as a measure of anxiety-proneness. *J. clin. Psychol.*, 1952, **8**, 279-283.
7. Harrower, M. R., & Steiner, M. E. *Large-scale Rorschach techniques.* (2nd Ed.) Springfield, Ill.: Charles C Thomas, 1951.
8. Hedlund, J. L., Farber, I. E., & Bechtoldt, H. P. Normative characteristics of the Manifest Anxiety Scale. *Amer. Psychologist*, 1951, **6**, 504. (Abstract)
9. Heineman, C. E. A forced-choice form of the Taylor Anxiety Scale. *J. consult. Psychol.*, 1953, **17**, 447-454.
10. Hilgard, E. R., Jones, L. V., & Kaplan, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94-99.
11. Holtzman, W. H., Calvin, A. D., & Bitterman, M. E. New evidence for the validity of Taylor's Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1952, **47**, 853-855.
12. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
13. Hurley, J. R. Verbal learning as a function of differentially motivating instructions and the achievement motive. Unpublished doctor's dissertation, State Univer. of Iowa, 1953.
14. Lauterbach, C. An empirical study of the Manifest Anxiety Scale and its relationships to other clinical measures of anxiety. Unpublished doctor's dissertation, State Univer. of Iowa, 1952.
15. Lucas, J. D. The interactive effects of anxiety, failure and intraserial duplication. *Amer. J. Psychol.*, 1952, **56**, 54-66.
16. Meyer, D. R., Bahrick. H. P., & Fitts, P. M.

Incentive, anxiety, and the human blink rate. *J. exp. Psychol.,* 1953, **45**, 183-187.

17. Montague, E. K. Role of anxiety in serial rote learning. *J. exp. Psychol.,* 1953, **45**, 91-96.
18. Myers, T. Discriminatory reaction performance as a function of test-defined anxiety level and verbal instructions. Unpublished master's thesis, State Univer. of Iowa, 1951.
19. Peck, Ruth. The influence of anxiety factors upon the effectiveness of an experimental "counseling" session. Unpublished doctor's dissertation, State Univer. of Iowa, 1951.
20. Rosenbaum, G. Stimulus generalization as a function of clinical and experimentally induced anxiety. Unpublished doctor's dissertation, State Univer. of Iowa, 1950.
21. Spence, K. W., & Farber, I. E. Conditioning and extinction as a function of anxiety. *J. exp.* Psychol., 1953, **45**, 116-119.
22. Spence, K. W., & Taylor, Janet A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.,* 1951, **42**, 183-188.
23. Taffel, C. Conditioning of verbal behavior in an institutionalized population and its relation to anxiety level. Unpublished doctor's dissertation, Indiana Univer., 1952.
24. Taylor, Janet A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.,* 1951, **41**, 81-92.
25. Taylor, Janet A. A personality test for manifest anxiety. *J. abnorm. soc. Psychol.,* 1953, **48**, 285-290.
26. Taylor, Janet A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.,* 1952, **44**, 61-66.
27. Wenar, C. Reaction time as a function of manifest anxiety and stimulus intensity. Unpublished doctor's dissertation, State Univer. of Iowa, 1951.
28. Wesley, Elizabeth L. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.,* 1953, **48**, 129-134.
29. Westrope, Martha. Relations among Rorschach indices, manifest anxiety, and performance under stress. *J. abnorm. soc. Psychol.,* 1953, **48**, 515-524.

Journal of Consulting Psychology
Vol. 18, No. 1, 1954

Mosaic Productions in Cerebro-arteriosclerosis[1]

Brendan A. Maher[2] and Anthony W. Martin[3]
College of Medicine, University of Illinois

Certain patterns of response in Lowenfeld's Mosaic Test have been described as characteristic of organically impaired subjects. Wertham [7], with regard to adults, and Colm [1] on the basis of work with children, report what they claim to be typical signs of organicity as revealed by Mosaic productions. Colm lists four criteria as follows: (*a*) the additive placement of pieces side by side, (*b*) the lining up of pieces of the same shape or color, (*c*) the piling of pieces vertically, and (*d*) the repetition of simple geometric designs.

Wertham describes two general categories of organic impairment to which he applies the terms "cortical" and "subcortical." The Mosaic pattern which is typically produced by the patient with cortically located brain damage is marked, he writes, "by an inability to achieve good configuration, with a dissolution and dismemberment of the Gestalt." The subcortical pattern is characterized by "stone-bound" designs, where at the "expense of an inner plan the patient follows the impetus inherent in the shape and color of the pieces put down so that the whole response becomes reduced to a mechanical or automatic level." In addition he provides one diagrammatic illustration [7, p. 248].

Elsewhere the Mosaic Test has been described as refractory to statistical analysis. A survey of the literature such as that recently completed by Dorken [2] indicates that the statistical testing of diagnostic hypotheses in this test has not yet been undertaken. Normative data have been provided by McCulloch and Girdner [4], Reiman [5], and Stewart and Leland [7] covering various aspects of the finished designs. However, so far as can be ascertained no hypotheses have as yet been clearly formulated and tested. Kerr [3] has reported a validity study based upon the matchings of Mosaics with over-all personality descriptions but gives no indication as to the nature of the interpretive principles used.

The study reported in this paper represents an attempt to define objective scoring criteria for some aspects of the Mosaic Test and to examine their validity. In order to avoid the ambiguities of definition involved in the study of psychiatric clinical groups, it was decided first to investigate the validity of the "organic signs" referred to above. In particular it was decided to examine the pattern characteristics described by Wertham. Organic impairment accompanying cerebro-arteriosclerosis was selected for study as this is regarded by Wertham as falling into both the cortical and subcortical categories.

In pursuit of this aim it was necessary to establish such accurately defined criteria that a trained observer might readily decide whether a given Mosaic production fell into the organic pattern or not. As written, Wertham's definitions suffer from some vagueness. The criteria used in this study were therefore stated more specifically as deductions from Wertham's broader terms and based upon his examples. The most striking aspect of the cortical pattern is the failure to achieve a good configuration. The question of what is "good" or "bad" in configurations is predominantly a matter of judgment on the part of the observer and therefore subject to question. However, it is possible to obtain data as to whether or not a patient has, in his own opinion, successfully completed a configuration. It was felt that the patient's report as to his perceived success or failure to complete a configuration offered the best hope of objectifying this principle. This was so defined under criterion 2 in the five criteria mentioned below.

---

[1] The writers wish to thank Dr. David Shakow and Dr. Alan K. Rosenwald of the Illinois Neuropsychiatric Institute and Mr. B. Grossman of the Drexel Home for their advice and assistance in this study.

[2] Now at The Ohio State University.

[3] Now at Purdue University.

With regard to the subcortical pattern, it can be seen that Wertham's concept of "stone-bound" or stimulus-bound is essentially descriptive of a species of perseverative response. However, it is possible for the subject to be bound to the characteristics of a stimulus in several ways. He may simply keep on placing pieces of the same shape, or of the same color, or both, in direct contact with each other. Alternatively he may try to produce over-all a Mosaic which has the same characteristics of form or color as its constituent pieces. Thirdly, he may perseverate at the level of repeating the process of putting pieces onto the tray with no discernible pattern and no continuity of structure. Criteria 1, 3, and 4 below were formulated to categorize these three possibilities. Experience with the test also suggested that it would be essential to exclude "representative" Mosaics, even though they might structurally meet the other criteria. By the term "representative" is meant any Mosaic which is intended to depict a concrete object, alphabetical or numerical symbol. Exclusion of this type of production was established as criterion 5. With this general rationale in mind the criteria were defined and used as a basis for the judges' selection as described in the next paragraph.

Method

The writers administered the Mosaic Test to 54 residents of a home for the aged in the city of Chicago. A record was made at the time of administration of all verbal comment, spontaneous or otherwise. The method of administration was the customary one wherein the subject is invited to make anything he wishes with the pieces on the tray. On completion a brief inquiry was conducted to determine what the subject had intended to make and whether he felt that he had been successful in his intention. The age of the group ranged from 64 to 91 years with a mean of 78.6 years. All the subjects were volunteers who responded to a request for individuals to cooperate in a study of aging. For record purposes the completed Mosaics were outlined and colored on the paper base. The records from the total group were then submitted to each of five psychologists on the staff of the Illinois Neuropsychiatric Institute. Each judge was asked to sort the completed productions into two groups on the basis of the five criteria: one group to be those which met one or another of the criteria, and the other to comprise those productions which did not meet any of the criteria. The criteria, as indicated earlier, were:

1. All designs or productions which consist of one or more single lines whether regular or irregular as to shape or color of pieces used.
2. All designs where the subject has failed in completing to his own satisfaction the design which he had stated he intended to make. This judgment involves reference to the verbal record sheet.
3. All designs which are "homogeneous" as to the relation of the shape of the pieces used to the final production; e.g., a square made of squares.
4. All designs in which pieces are simply placed side by side with no discernible over-all pattern, irrespective of shape or color of pieces used.
5. A negative criterion. No successfully completed design which is of a representational character is to be included, even though it may fulfill one of the criteria. Thus a "square made of squares" would not be included if the subject had described it as a "tile floor" and had completed it to his own satisfaction.

Of the 54 productions judged, a total of 27 were selected by the judges as meeting one or another of the criteria. Of this number, 20 were chosen by all five judges. The seven ambiguous cases were discarded from the sample leaving a further 27 cases which had been unanimously judged as meeting none of the criteria. There was therefore complete agreement on 87 per cent of the cases judged.

Following the classification of Mosaics, reference was made to the results of the routine physical examination which had been given to these subjects by the institution physician. These examinations had been made before and independently of this study. On the basis of these records the subjects were classified into those having recognizable signs of cerebral arteriosclerosis with deterioration and those not having such signs. Five of the cases in which the report was equivocal were discarded. Those subjects whose Mosaics had been discarded as ambiguous were also excluded, leaving ultimately a total of 46 subjects, 20 of whom showed signs of cerebral arteriosclerosis. From the remaining 26, a further 20 having good health and most easily matched in mean age with the organic group were selected. The matching achieved was as follows: The mean age of the organic group was 75.9 years, with *SD* of 7.3 years, and of the control group, 76.5 years, with *SD* of 6.8 years.

The groupings made by the judges were then analyzed with respect to the frequency of distribution of "organic" patterns in both groups. The "organic" Mosaic productions were then further subdivided into those judged as meeting criteria 1, 3, or 4, (the subcortical criteria) and those meeting criterion 2 (the cortical criterion). The results are given in Table 1.

Table 1

Distribution of Mosaic Patterns in a Group of Aged

Group	A Criteria 1, 3, and 4	B Criterion 2	C No Criteria
Organic	10	7	3
Control	2	1	17
χ^2	11.7	10.11	
vs. col C*	$p = .001$	$p = .01$	

* Chi square calculated using Yates's correction.

Discussion

The levels of significance obtained support the inference that there is a meaningful relationship between the presence of cerebral arteriosclerosis, as judged by physical examination, and the production of certain patterns on the Mosaic Test. In general the data support the clinically derived interpretive principles in so far as the criteria used in this study can be regarded as legitimate inferences from Wertham's rather broader definitions. However it should be pointed out that the criteria are not all derived from the same level of description of psychological events. Criterion 2 differs from the rest in that it is not confined to a description of the structural features of the finished Mosaic production but deals rather with a failure in planning as reported by the subject himself. Influences covered by the psychological concepts of level of aspiration, anticipation, and perhaps self-concept may be involved here. The relationship of organic disability to changes in these influences remains undisclosed by a study such as this. Nevertheless the clinician concerned with diagnostic problems rather than theoretical issues may find criteria drawn from several levels of description to be useful aids in his work. It is to this end that this report is primarily directed. The more fundamental implications for theories of perception and projection remain problems for further investigation.

Received May 1, 1953.

References

1. Colm, Hanna. The value of projective techniques in the psychological examination of children: the Mosaic Test in conjunction with the Rorschach and Binet tests. *Rorschach Res. Exch.*, 1948, **12**, 216-233.
2. Dorken, H. The Mosaic Test: review. *J. proj. Tech.*, 1952, **16**, 287-296.
3. Kerr, M. The validity of the Mosaic Test. *Amer. J. Orthopsychiat.*, 1939, **9**, 232-236.
4. McCulloch, T. L., & Girdner, J. B. The use of the Lowenfeld Mosaic Test with mental defectives. *Amer. J. ment. Def.*, 1949, **53**, 486-496.
5. Reiman, M. G. The Mosaic Test: its applicability and validity. *Amer. J. Orthopsychiat.*, 1950, **20**, 600-615.
6. Stewart, U., & Leland, L. American versus English mosaics. *J. proj. Tech.*, 1952, **16**, 246-248.
7. Wertham, F. The Mosaic Test: technique and psychopathological deductions. In L. Abt, & L. Bellak (Eds.), *Projective psychology.* New York: Knopf, 1950. Pp. 230-256.

Journal of Consulting Psychology
Vol. 18, No. 1, 1954

Sexual Symbolism: A Validity Study

Leon H. Levy
The Ohio State University

Dating from Freud's work on dream analysis [1], and perhaps before, there has been present in psychology and in psychoanalysis, in particular, the working hypothesis that certain environmental objects stand in a symbolic relation to the male and female genitalia.

As originally elaborated by Freud, the symbolic relationship was based on certain structural or functional similarities between the symbol and the genital organ symbolized. Thus, elongated and pointed objects, such as knives, telephone poles, chimneys, and noses were assumed to represent the male genitalia, while round, containing objects, such as boxes, key holes, houses, and ships were considered symbolic of the female genitalia.

Although White [6] has claimed that this hypothesis is no longer in good repute, it would seem instead that it has gained a wide, implicit acceptance and has now become, for many, an integral part of their interpretation of projective materials ranging from figure drawings [5] to the Bender-Gestalt [4].

This is not to say that stereotyped interpretations are necessarily made, or that the productions are considered out of context. In one case the absence of a chimney on a house, for example, may be said to indicate that the individual feels inferior to his father in terms of power and authority while in another this may be taken as a sign of feelings of sexual impotence. In either case, however, the object has been assumed to have a certain fixed symbolic relationship with reality, i.e., chimney = penis.

The purpose of the present study is to investigate this hypothesized relationship. The hypothesis to be tested may be stated as follows: Where abstract objects are categorized by a given means as either male or female, those objects which may be described as elongated, pointed, or angular will be categorized as male; those objects which may be described as round and containing in nature will be categorized as female.

Method

The experimental design was developed with two factors in mind. The first was that the research was being carried on in a public school system where protest and resistance might be encountered should the purpose of the research be known. Secondly, from the standpoint of validity, it was felt that an indirect method of a projective nature would be less subject to any conscious resistances or anxieties which the subjects (*S*s) might have concerning sex. Further, by using an indirect method it was hoped that the hypothesis was being tested at the level at which it is assumed to be operating.

The *S*s consisted of 62 children, 30 male and 32 female, in the fifth grade of a public school in a community of 100,000 population. The *S*s were divided into two groups, A and B, of 35 and 27, respectively. There was an equal proportion of males and females in each group.

The hypothesis was tested by two approaches, each designed to operate at a different level of awareness. In the first approach, *S*s in both groups were told that the task they were about to perform was part of an investigation of immediate memory. The *S*s were presented with a sheet of paper on which ten first names were randomly listed, five of each sex. Below the list of names, ten figures were presented in random order. Five of the figures were elongated or pointed and were therefore assumed to be symbolic of the male genitalia; five were rounded and containing or symbolic of the female genitalia.[1] The experiment was divided into three parts. During the first part

---

[1] Copies of the figures are available on request from the author.

Table 1

Predicted and Observed Frequency Distributions for Any Given Number of Correct Matches

Number of correct matches	p	Predicted frequency	Observed frequency
0	.001	.2	0
2	.10	6.2	2
4	.39	24.6	29
6	.39	24.6	23
8	.10	6.2	8
10	.001	.2	0

*S*s were asked to assign one of the ten names to each of the figures, using each name only once. The sheets were then collected and *S*s not adhering to the instructions were eliminated from the study.

The second part consisted of paired-associates learning. Ten flashcards each containing one of the ten figures and one of the names were exposed for three seconds each in immediate succession. For Group A, the elongated and pointed figures were presented with male names, while the rounded and containing figures were shown with female names. For Group B, figures symbolic of the male genitalia were presented with female names, and the female figures with male names.

Part three consisted of a retention test in which *S*s from Groups A and B were asked, on the basis of the flashcards, once more to match symbols with names. Matching was done on blank sheets like those used in part one.

The hypothesis being tested may now be stated in terms of the following prediction: 1. In part one, matching of figures and names so that they are both of the same "sex" would be greater for the entire sample than that expected by chance. 2. On the retention test in part three, Group A would show better retention than would Group B.

The first prediction is based on the assumption that in any forced matching situation, those items perceived as related in any fashion will be matched more often than will items where there is no perceived relation if there are no instructions to the contrary. If the hypothesis being tested is valid, there should be certain regularly perceived relationships between the figures and the names.

In Bartlett's terminology [1], one may assume that if the hypothesis is valid, there exists for the *S*s a schema whereby there is a definite relationship between type of figure and sexuality. Thus, in part two the learning of Group A may be considered as *pro-schema* learning and that of Group B, *anti-schema* learning. The experimental design as a whole may be viewed as a study in negative and positive transfer effects; hence the second prediction based on the assumption that Group A will be subject to positive transfer effects while Group B will be subject to negative transfer effects.

Thus, while *S*s may not associate names and figures in part one as predicted owing to certain conscious inhibitions, if this association does exist it should affect the learning of the two groups in part two.

Table 2

Comparison of Groups on Retention Scores

Group	N	Mean	SD	t
A	35	5.7	1.907	1.384
B	27	5.111	1.929	$>.05$

Results

Since the first part of the experiment represents a simple sorting situation, it is possible to test the null hypothesis by first calculating the probability of getting any given number of "correct" sorts, translating this into an expected frequency distribution for the experimental sample, and comparing this by means of chi square with the actual frequency distribution. The results are presented in Table 1.

Chi square for this table is 4.714 which with 7 degrees of freedom is not significant and therefore gives no reason for rejecting the null hypothesis.[2] Hence, the first prediction derived from the experimental hypothesis is not validated.

Table 2 presents the data on scores obtained on the retention test in phase three. The difference between the means of the recall scores for the two groups does not reach statistical significance and hence fails to confirm the second prediction derived from the hypothesis.

---

[2] In calculating χ^2 the two extreme categories at either end were combined because of the low predicted frequencies [3].

While the mean recall scores were not significantly different for the two groups, there would still be some support for the hypothesis if the types of errors on recall made by the two groups were different, i.e., if the errors of Group B could be shown to be due to intrusions of previously determined associations between type of figure and sex. An intrusion for Group B would be any error where the *S* has matched a male figure with a male name and female figure with female name. In other words, these are errors counter to training. In order to evaluate the significance of such intrusions for Group B, it seems appropriate to compare the percentage of their total errors due to such intrusions with the percentage of errors made by Group A which were counter to training or "intrusions." This is presented in Table 3. It will be noted that there is no difference between these two percentages.

Table 3
Analysis of Errors Made on Retention Test

Group	Total number of errors	Number of errors due to intrusions	% of total errors
A	147	56	38
B	132	50	38

Discussion

The results of this experiment do not lend any support to the hypothesis of sexual symbolism as stated at the outset. While it is true that the results of this investigation should not be generalized much beyond its experimental population, its results do have broader significance in so far as the proponents of universal sexual symbolism do not set any limits on its applicability.

Since man is a symbol-using animal, it is quite reasonable to expect him to develop many symbols related to sexuality. However, the results of this experiment would throw into serious question the universality of his symbolism. Blind analysis of projective materials, which has become the current fashion, would certainly appear to be on tenuous grounds in so far as it relies on hypotheses similar to the one which was tested here.

The face validity and tenaciousness of the hypothesis of universal sexual symbolism may be explained by several factors. The most important perhaps is that the population out of which this hypothesis was derived was a treatment population for the most part. As such, it might be expected, first, that they would be more likely to have sexual problems and to utilize symbolic representation than would the general population, and second, should the therapist have certain preconceived notions concerning symbolism, it is quite likely that these individuals would be more susceptible to suggestion and hence provide the therapist with the validation he seeks.

Further, because of the uncontrolled nature of the observations made by therapists as well as the likelihood of selective perception, it is quite likely that this is another example of the operation of partial reinforcement.

Summary

The present investigation was aimed at testing the hypothesis that certain environmental objects by virtue of their structural characteristics may be taken as being symbolic of the male and female genitalia. The *S*s used were normal children in the fifth grade of a public school.

The hypothesis was tested by means of a simple matching situation and a learning situation, both of which involved male and female first names and abstract figures conforming structurally to either the male or female symbol as described in the literature. In the first part of the experiment, *S*s were asked to match names and figures, and in the second part they had to learn to associate certain names with certain figures. The latter represented essentially a test of positive and negative transfer effects as would have been predicted by the hypothesis.

The results of both parts of the experiment failed to lend support to the hypothesis. The results were interpreted as evidence against the existence of a universal sexual symbolism and an indication of the need for extreme caution in the interpretation of symbolism and in blind analysis.

Received April 27, 1953.

References

1. Bartlett, F. C. *Remembering.* London: Cambridge Univer. Press, 1932.
2. Freud, S. *The basic writings of Sigmund Freud.* New York: Modern Library, 1938.
3. Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1942.
4. Hutt, M. L. *A tentative guide for the administration and interpretation of the Bender-Gestalt Test.* U. S. Army, Adjutant General's School, 1945 (restricted).
5. Vernier, C. M. *Projective test productions. I. Projective drawings.* New York: Grune & Stratton, 1952.
6. White, R. W. *The abnormal personality.* New York: Ronald, 1948.

Journal of Consulting Psychology
Vol. 18, No. 1, 1954

The Interpretations of Anxiety by Various Psychotherapeutic Schools[1]

John H. De Michele
University of Illinois

Contemporary need for a briefer and more effective psychotherapy has led to many searching revisions of methodology. This trend can be seen in the works of Alexander and French [1], Horney [11], Mowrer [15], Rogers [18, 19], Thorne [21], and others. Valuable as this research is, the intensive study by each school of its particular techniques is a narrow approach to methodology. Since each school claims excellent results, it is possible that there may be common factors leading to this therapeutic success. An analysis by White [22], for example, found that although there are many psychotherapeutic methods, there are only five processes or techniques common to all schools. In order to test this conception, studies comparing the various psychotherapeutic schools on these common basic techniques are needed and might prove fruitful. An encouraging beginning has been made in the realm of the patient-therapist relationship by Fiedler [7], Heine [10], Miller [14], and Quinn [17]. This patient-therapist relationship would approximate White's first and fourth process, i.e., rapport and transference. Fiedler, for example, found an optimal patient-therapist relationship basic and common to many schools of psychotherapy. Another basic method is interpretation. Although studies comparing the relative effectiveness of the various forms of psychotherapeutic interpretation have been made by Grossman [8] and a few earlier investigators [5, 9, 12], no studies have been made comparing the various psychotherapeutic schools on the technique of interpretation per se. The basic technique of interpretation is essentially White's third process, called the *pointing out of feelings*. This process has two forms: (*a*) the Rogerian *recognition of feeling* and (*b*) *interpretation,* which categorical as they are, White states, "are not sharply different from each other" [22, p. 345]. Now, Bordin [2], Collier [6], and Stone [20] have suggested the very important point that we think in terms of a continuum along any basic method, i.e., that each therapist varies in the amount of development that he allows on the continuum of any basic method. Actually, there are many continua which can be found within any broad basic method. Collier, for example, proposes a continuum of "depth" of interpretation. The idea of continua is valuable, since it can lead to the possibility of measuring variations within each of the broad basic methods. The comparative study of various schools of psychotherapy is valuable in that it can lead to the discovery of continua and the experimental testing of basic psychotherapeutic methods.

In this study, interpretation has been selected for analysis. Interpretation may be defined as the inference by a therapist of the attitudes of a patient from cues supplied by facial expression, posture, language, gesture, dreams, play, etc. The broad objective of the experiment will be to discover continua within the basic method of interpretation; to discover quantitative variations among the different methodological groups of therapists in terms of *what* is selected to substantiate an inference, and *how* they arrive at an interpretation.

The Criterion

In order to carry out the broad objective of this study, the criterion selected for interpretation should be as definitive as possible. An intense degree of anxiety was selected as the criterion, because its somatic manifestations are directly observable physiological reactions. The term *anxiety* is used to represent all of its forms, i.e., anxiety neurosis, prodromal anxiety, normal anxiety, or fear, since the symptoms are the same. While differences between these forms may be inferred or observed situationally, internally, or in terms of feelings, Brown [3, p. 131] and May [13, pp. 190, 203] agree that the directly observable symptoms are the same. A subsidiary objective of the experiment will be to determine which of the most common directly observable symptoms are used to substantiate an inferred intense degree of anxiety.

---

[1] This paper is an outgrowth of the thinking of Dr. L. I. O'Kelly, to whom I am grateful for constant stimulation and encouragement.

From the most complete list [4, p. 48] of anxiety symptoms available, a selection of directly observable symptoms was made. To these were added difficulty in sleeping, because its effects can be noticed, and fear, because its intimate relationship should have the highest correlation with anxiety. The symptoms selected were as follows:

1. Jumpiness to slight sounds, lights, etc.
2. Great tension.
3. Agitation to the point of trembling.
4. Great difficulty in breathing.
5. Profuse sweating.
6. Sleeps very poorly.
7. Great fear.
8. Great difficulty in concentration.
9. Great difficulty in speaking.

The Media

Since the symptoms of anxiety are directly observable, and since the inference of anxiety is essentially the judgment of an emotional expression, 10 pictures of people showing various degrees of emotional expression were selected as vehicles which might be conducive to inferring anxiety. The pictures ranged from nondescriptive expressions to intense expressions suggestive of fear, anxiety, or terror. The pictures with the intense emotional expressions of fear were selected on the basis of similarity to the typical ones used in earlier studies concerned with the judgment of emotion [16, pp. 365-369]. Each picture was of one person in a relatively vague background, so that little could be inferred from the situational context.

The Preliminary Experiment

A preliminary experiment was conducted for two purposes: (*a*) to see if judgment of the symptoms of anxiety has a good correlation with a judgment or inference of anxiety, and (*b*) to see if the expected range of emotional expression, subjectively selected by the criteria of facial and postural expressions of emotion, is as diverse as the measured range of the emotional expression suggested by the pictures.

A group of 17 psychologists, composed of clinicians from the Veterans Administration Hospital at Downey, Illinois, and counselors from the Chicago Undergraduate Division of the University of Illinois, was selected for the preliminary experiment.

Each psychologist rated the selected symptoms of anxiety arranged in five-point scales for each of the 10 pictures shown. He then had to rate anxiety and two other clinical syndromes on similar five-point scales. The median rating was found for each of the 10 pictures, and each was correlated by Spearman's rho with the median anxiety ratings for each of the 10 pictures. The correlation coefficients for each symptom with anxiety ranged from .35 to .89. The mean of all of the median ratings was found and correlated .87 with anxiety. We can conclude from the evidence that clinicians who rate these symptoms high tend to rate anxiety high.

The only quantitative indication that the emotional expressions of the persons shown in the pictures are well dispersed is the median score of the amount of anxiety seen by the clinicians in each picture. Pictures 2, 6, and 10 were always judged to have a low degree of anxiety, while picture 1 was always judged to have a high component. Pictures 4 and 9 were judged to have a high degree of anxiety two out of three times, while pictures 3, 5, 7, and 8 varied within the middle area of the scale.

Methodological Types of Therapists

Two distinct methodological schools may be identified as Rogerian psychotherapy and Freudian psychonanalysis. Adherents of these schools should show the greatest methodological consistency. The "eclectic" psychologists and psychiatrists, although each has a distinct training, are less homogeneous. The groups used in this study consisted of 10 Rogerian psychotherapists from the Counseling Center of the University of Chicago, 10 psychoanalysts who were trained at the Chicago Institute of

Psychoanalysis, 6 eclectic psychologists, and 5 eclectic psychiatrists. An effort was made to select the most efficient and experienced therapists; five years of therapeutic experience was generally required.

Procedure

The pictures showed a person expressing some degree of a feeling, just as a person might in a face-to-face therapeutic situation. In section 1 of the directions, the therapist was asked to infer what he thought the person in the picture was feeling. In section 2, the therapist was then asked what in the picture led him to his impression; he was further asked what other signs, cues, or symptoms he would *expect* the person to display. In the case of anxiety this procedure led to a selection of the most diagnostic signs that each therapist uses to infer anxiety. Section 3 of the directions consisted of rating the signs in terms of the magnitude of their contribution to the major impression, but this rating was not too successful. After all of the pictures had been completed, section 4 of the directions was presented; here the therapists had to classify each of the reshown pictures into one of four affective states: (*a*) normal, (*b*) neurotic, (*c*) psychotic, and (*d*) anxiety. In addition, he had to rate each of these states in terms of intensity, from 1 to 4. Section 4 was of value in two ways: first, to check the consistency of the therapist. If he saw anxiety in the initial inference he should also see it in the final classification; and second, to check the relative intensity of the anxiety classification.

The order of presentation was maintained constant, and the therapist had to comply with at least sections 1 and 4 of the directions. Each therapist had to write his own test. The time given to the administrator by the therapist was of major importance in the flexible use of the other two sections of the directions. Under pressure of time, only the material directly relating to anxiety was completed. The therapists were especially urged to do their best when they inferred anxiety.

Qualitative Results and Observations

Most of the therapists were cooperative, interested in the test, and readily made available their facilities. Many psychoanalysts, however, were rather defensive. The writer found many interesting reactions. Some therapists were efficient and quickly understood the directions; others had to have the directions read over and over again. Two therapists apparently experienced anxiety reactions when they could not cope with a picture. These therapists could not go on with the test and appeared very agitated; one perspired freely, while the other had to leave the room for a smoke to settle his nerves. Some therapists thought the test was difficult, others thought it simple. As a whole the therapists disliked the idea of writing the test and would have preferred a recording machine. The protocols of many therapists showed conflicts, and some stated that the test was a "personal" experience for them.

A sample of an excellent handling of the criterion in picture 1 by an eclectic psychologist is shown in the following protocol: The woman is distressed and frightened by something which she sees happening at her right. She is clutching her coat (1), has a puckered brow, and a tense anguished expression about the mouth (1). If one could see this woman through time, i.e., in motion, she might be restless or more or less frozen. One might expect *explosive* behavior, a sudden decision to do some particular thing. She would be inclined to vocalize, breathing would almost certainly be disturbed (4); there might be anomalies of perspiration, e.g., (cold perspiration) (3); there might be a high degree of concentration, preoccupation (2). If someone spoke to her she would either not hear them, perhaps be petulant, or turn eagerly to them. (D-3)

In order to typify the group's approach, the following examples of picture 1 are presented.

A Rogerian therapist's protocol, typical of the group, is as follows: "I think she is in real emotional distress, with fear and pain as her prominent feelings. Cues: the knotted brow (2), the clenched hand (1), muscular strain on the face (3). I would expect weeping, very likely sobbing to follow if she started to explore her feelings." (D-4)

A typical eclectic psychologist's protocol is as follows: "A reaction of panic, consisting of fright, sorrow and anxiety. She is immobilized as far as reacting is concerned (4). Facial expression (2) and position of the hand (1); general impression that she is viewing something in the distance (1). She feels numb and cold and at the same time agitated (1). Her face is hot and she is perspiring. (1) Her mouth is dry and her heart is pounding. (1) Tears are about to come." (D-4)

A typical analyst's protocol is as follows: "Apparently a frightened woman — looks unhappy — and some pain I feel from internal sources; (4)

the painful expression of the eyes, (3) hand clutched to the coat, (1) mouth open. She seems to be looking at something which gives this reaction of fright — she might next turn away. However I've seen psychotic patients on walks react like this from no external stimuli. Accounting for my first impression above. Also there is a pseudo-quality or an overreaction of pain which vanished quickly and the patients appeared more relaxed following it. She may cry from this point on; (1) lower her head, (2) turn her head, (3) show beads of perspiration." (D-4)

Quantitative Results

Inference from subjective feelings. This factor appears to lie on a continuum which has crude quantitative verification. When the test was being administered to analysts, the writer became aware of an involved "soul searching" to find cues to substantiate their inferences. On the other hand, the Rogerians and eclectic psychiatrists overtly seemed little involved. Some therapists felt so strongly about this internal source of cues that they had to write them down. When a therapist is so aware of his feelings as cues, we could expect a different kind of interpretation than from a therapist who objectifies the problem. The protocols show four analysts, two eclectic psychologists, one Rogerian, and no eclectic psychiatrists who used these internal sources as cues. Crude as this evidence may be, it certainly should be investigated further.

Inference of pathology. The results from the classification of affective states under section 4 of the directions was arbitrary and in many cases unsatisfactory; yet a scrutiny of the results showed a wide difference in classification by the different groups. A dichotomy of the classifications was made between the normal and the three abnormal affective states. The analysts inferred the least amount of normality in the set of pictures, and conversely the greatest amount of pathology. The Rogerians and the eclectic psychiatrists inferred the least amount of pathology, while the eclectic psychologists lie in between the two poles. The significance of differences between the mean amount of pathology inferred by the psychotherapeutic groups was found with the Mann-Whitney *U* Test. The results can be seen in Table 1.

Table 1
Significance of Differences between Psychotherapeutic Groups on the Inference of Pathology
(Using the Mann-Whitney *U* Test)

Groups	*N*	*U*, *U–z* score		*p*
1. Rogerians and analysts	19	(*U–z*)	2.82	.005
2. Eclectic psychiatrists and analysts	15	(*U–z*)	2.82	.005
3. Eclectic psychologists and Rogerians	15	(*U–z*)	1.97	.05
4. Eclectic psychologists and eclectic psychiatrists	11	(*U*)	5	.08
5. Eclectic psychologists and analysts	16	(*U–z*)	1.29	NS
6. Eclectic psychiatrists and Rogerians	14	(*U–z*)	.60	NS

The inference-cue ratio. The inference-cue ratio is the number of inferences made as compared to the number of cues used to support these inferences. For example, if a therapist makes four inferences and uses only two cues to support his inferences, then his inference-cue ratio would be four divided by two, or two. Two inference-cue ratios were found. The first ratio was the number of inferences made under section 1 of the directions to the number of cues selected from the picture. For the second ratio, the expected or inferred cues, signs, or symptoms were treated as inferences; thus, this ratio was the number of inferred cues to the number of cues selected or observed in the picture. The inference-cue ratios were all made on picture 1, which was completed by all therapists, and upon which the directions were best controlled. Table 2 shows the in-

Table 2
The Inference–Cue Ratios of the Psychotherapeutic Groups

Therapists	Rogerians		Eclectic psychiatrists		Eclectic psychologists		Analysts	
	I/C*	IC/C*	I/C	IC/C	I/C	IC/C	I/C	IC/C
1.	0.5	0.0	0.8	1.8	2.5	4.0	1.8	0.0
2.	1.0	0.8	1.5	1.1	1.0	0.8	2.0	3.0
3.	1.2	0.0	1.0	1.7	1.1	1.0	0.5	1.5
4.	0.8	0.0	1.1	0.0	1.0	4.0	1.0	2.3
5.	0.6	0.6	1.0	2.0	0.8	0.0	1.3	2.0
6.	1.0	0.6			0.7	2.0	2.0	2.0
7.	1.0	0.0					1.5	1.1
8.	1.0	0.0					2.7	1.3
9.	2.3	1.6					4.0	3.0

* I — inferences, IC — inferred cues, C — cues.

Table 3
Significance of Differences between Psychotherapeutic Groups on the Inference-Cue Ratios (Using the Mann–Whitney *U* Test)

Groups	*N*	*U*, *U*–*z* score	*p*
1. Rogerians and analysts	18		
I/C ratio*		(*U*–*z*) 1.94	.05
IC/C ratio*		(*U*–*z*) 2.74	.006
2. Eclectic psychiatrists and analysts	14		
I/C ratio		(*U*–*z*) 1.58	NS
IC/C ratio		(*U*–*z*) 1.03	NS
3. Eclectic psychiatrists and Rogerians	14		
I/C ratio		(*U*–*z*) .66	NS
IC/C ratio		(*U*–*z*) 2.05	.04
4. Eclectic psychologists and Rogerians	15		
I/C ratio		(*U*–*z*) .35	NS
IC/C ratio		(*U*–*z*) 2.13	.03
5. Eclectic psychologists and eclectic psychiatrists	11		
I/C ratio		(*U*) 13	.08
IC/C ratio		(*U*) 17	.08
6. Eclectic psychologists and analysts	15		
I/C ratio		(*U*–*z*) 1.53	NS
IC/C ratio		(*U*–*z*) .17	NS

* I — inferences, IC — inferred cues, C — cues.

ference-cue ratios of the various psychotherapeutic groups.

From the inference-cue ratios of Table 2, we can observe a progression from the Rogerians, who clearly show an extreme reluctance to infer or interpret at all, to the analysts, who interpret most from the cues they selected from the picture. This progression is an indication of the "depth" of interpretation continuum. The significance of difference of the two inference-cue ratios was found between the various psychotherapeutic groups; the results can be seen in Table 3.

Nosological orientation. For each group of therapists, the number of consistent inferences of anxiety was found, i.e., where anxiety or fear was inferred in section 1 of the directions and confirmed in section 4 with a rating of D-3 or D-4. Ten analysts gave 21 consistent inferences of anxiety throughout the 10 pictures, while nine Rogerians at the other end gave 8 consistent inferences of anxiety. The ratio of the number of times each symptom is used compared to the number of consistent anxiety inferences made by each group may be expressed as a percentage. This percentage shows how often a particular symptom is used to support a consistent inference of anxiety. In addition, Table 4 shows which symptoms are used most often by the various groups to substantiate their inferences of anxiety.

Table 4
Percentage of the Symptoms of Anxiety Used to Support a Consistent Inference of Anxiety

Symptoms	Analysts	Eclectic psychiatrists	Eclectic psychologists	Rogerians
1. Perspiration	57.1	71.4	33.3	0.0
2. Breathing difficulty	28.5	57.1	25	0.0
3. Immobility	14.3	14.3	33.3	0.0
4. Vasomotor difficulty	14.3	28.5	8.3	0.0
5. Tension	9.5	85.5	33.3	0.0
6. Pupils dilated	9.5	28.5	0.0	0.0
7. Confusion	4.8	14.3	8.3	0.0
8. Exhaustion	14.3	14.3	16.6	0.0
9. Rectal urgency	4.7	14.3	0.0	0.0
10. Pilomotor reaction	0.0	14.3	0.0	0.0
11. Speech difficulty	0.0	14.3	8.3	0.0
12. Vomiting	4.7	0.0	0.0	0.0
13. Nervous movements	9.5	14.3	0.0	0.0
14. Collapse	9.5	14.3	0.0	0.0
15. Flight	19.0	28.5	16.6	12.5
16. Cryout or tears	28.5	57.1	25.0	25.0
17. Gross body movements	47.6	0.0	0.0	62.5

The eclectic psychiatrist, whose training is heavily weighted with nosology, would probably substantiate an inference of anxiety with cues which are more nosologically oriented than would the Rogerian, who at the other extreme has little interest in nosology. A continuum in terms of nosological orientation has some experimental confirmation. There is a progression of the number of nosological symptoms used, extending from the analysts and eclectic psychiatrists at one end to the Rogerians at the other. None of the Rogerians inferred any of the nosological symptoms of anxiety, but instead they used gross bodily movements descriptive of the startle pattern. The inferred or expected cues used were crying, screaming, expressing concern, and movements away or toward the anxiety-producing object.

Summary and Conclusions

The aim of this research was an attempt to compare the various psychotherapeutic schools on the basic method of interpretation. Therapists were to interpret anxiety from pictures showing various degrees of emotional expression. The results were analyzed qualitatively and quantitatively. Four continua in the basic method of interpretation were postulated from the results of the experiment. These continua are tentative and need more refined experimental substantiation. The four continua are:

1. "Depth" of interpretation
2. Nosological orientation
3. Inference of pathology
4. Inference from subjective feelings

Received June 8, 1953.

References

1. Alexander, F., & French, T. M. *Psychoanalytic therapy.* New York: Ronald, 1946.
2. Bordin, E. S. Dimensions in the counseling process. *J. clin. Psychol.*, 1948, **4**, 240-244.
3. Brown, J. F. *The psychodynamics of abnormal behavior.* New York: McGraw-Hill, 1940.
4. Cattell, R. B. *Description and measurement of personality.* Yonkers, N. Y.: World Book Co., 1946.
5. Closson, F. H., & Hildreth, H. M. Experiment in psychotherapy during selection examination. *Nav. med. Bull.*, 1944, **43**, 39-43.
6. Collier, R. M. A basis for integration rather than fragmentation in psychotherapy. *J. consult. Psychol.*, 1950, **14**, 199-205.
7. Fiedler, F. E. The concept of the ideal therapeutic relationship. *J. consult. Psychol.*, 1950, **14**, 239-245.
8. Grossman, D. The experimental investigation of a psychotherapeutic method. *J. consult. Psychol.*, 1952, **16**, 325-331.
9. Haggard, E. A., & Murray, H. A. The relative effectiveness of three "therapy" procedures on the reduction of experimentally induced anxiety. *Psychol. Bull.*, 1942, **39**, 441. (Abstract)
10. Heine, R. W. An investigation of the relationship of change in personality from psychotherapy as reported by patients and the factors seen by the patients as producing the change. Unpublished doctor's dissertation, Univer. of Chicago, 1950.
11. Horney, Karen. *New ways in psychoanalysis.* New York: Norton, 1939.
12. Keet, C. D. Two verbal techniques in a miniature counseling situation. *Psychol. Monogr.*, 1948, **62**, No. 7 (Whole No. 294).
13. May, R. *The meaning of anxiety.* New York: Ronald, 1950.
14. Miller, Helen E. "Acceptance" and related attitudes as demonstrated in psychotherapeutic interviews. *J. clin. Psychol.*, 1949, **5**, 83-87.
15. Mowrer, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.
16. Munn, N. L. *Psychology.* Boston: Houghton Mifflin, 1951.
17. Quinn, R. D. Psychotherapists' expressions as an index to the quality of early therapeutic relationships established by representatives of nondirective, Adlerian, and psychoanalytic schools. Unpublished doctor's dissertation, Univer. of Chicago, 1950.
18. Rogers, C. R. *Counseling and psychotherapy.* Boston: Houghton Mifflin, 1942.
19. Rogers, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.
20. Stone, D. R. Logical analysis of the directive, non-directive counseling continuum. *Occupations,* 1950, **28**, 295-298.
21. Thorne, F. C. *Principles of personality counseling; an eclectic viewpoint.* Brandon, Vt.: Journal of Clinical Psychology, 1950.
22. White, R. W. *The abnormal personality.* New York: Ronald, 1948.

Journal of Consulting Psychology
Vol. 18, No. 1, 1954

The Relationships among Projection, Empathy, Reality, and Adjustment, Operationally Defined

Ralph D. Norman and Patricia Ainsworth
The University of New Mexico

The hypotheses of this study are two: (*a*) when operational definitions of projection, empathy, and reality are given, projection will be negatively correlated with both reality and empathy, whereas empathy and reality will be positively correlated with each other; and (*b*) reality and empathy will correlate more closely with adjustment than will projection.

Regarding the first hypothesis, Allport [2], Healy, Bronner, and Bowers [12], and Sears [15] state that projection can occur provided the individual lacks insight into himself. But self-insight requires a degree of realistic perception of oneself. Steinmetz [16] expands the concept of insight to mean the ability to perceive the world as others do. In either case, it may be stated, as does Alexander [1], that projection involves a lack of reality testing. He says, using projection of hostility as an example, "If this hostility becomes conscious it would cause guilt and anxiety, and to avoid this the whole situation may be distorted by saying, 'I do not want to attack him; he wants to attack me.' Such distortion of reality becomes possible only when the ego renounces its contact with reality under the pressure of emotional needs" [1, p. 114]. Thus, it may be postulated that a negative relationship exists between projection and realistic perception or reality.

That projection is also negatively related to empathy and that empathy is positively related to reality may be inferred from logic and previous research. Certainly, empathy, or the ability to put oneself in another's place, requires realism in seeing others. Dymond [6] has shown that insight and empathy are related to each other, and Norman [13] has demonstrated that there are positive relationships among variables he terms insight, self-other identity, and realistic perception of others. As part of the first hypothesis, then, it is the intention of this study to confirm these findings by Dymond and by Norman, using an entirely different technique. It may also be logically inferred that if reality and projection are negatively related, and if reality and empathy are positively related, then empathy and projection should be negatively related.

Regarding the second hypothesis, it may be predicted that adjustment should be more closely tied to empathy and reality than to projection. Projection is a defense mechanism of the ego and inherently implies maladjustment and insecurity, according to Bellak [3]. It is also prominent in paranoid maladjustment, as is well known. Moreover, Dymond [8] found that individuals who possess empathic ability are better adjusted and more secure than those who do not. Norman [13], using sociometric data of acceptance, found positive relationships between acceptance and self-other identity, insight into self, and realistic perception of others.

Subjects and Procedure

The Guilford-Martin Inventory of Factors GAMIN [11] was given to 74 male college students, most of whom were freshmen. This inventory was chosen because it is easy to administer, was standardized on students, and has high reliability. Particularly, the fact that the test has been factor analyzed is especially advantageous to this study because of interest in the operation of the various experimental variables (empathy, projection, and reality) with traits of high factorial validity. Guilford and Martin [11] state that the intercorrelations between the traits are generally low enough to assure useful differences in measurement between the trait scores. By giving the test in two forms, scores for our three experimental variables were obtained. The first form was administered as usual except that the "?" alternative in answers was omitted. The subjects thus had to answer either "Yes" or "No." Two weeks later, a second test was given to the same subjects, and, instead of answering for themselves, subjects were required to answer the questions as they thought most others of their own age and sex would answer them. For example, Question 8 on the original, "Do you feel that people almost always treat you right?" was changed to read, "Do you think that most people of your own age and sex feel that people almost always treat them right?" It was felt that there would be more uncertainty in answering questions for others and therefore that too many question marks would be used. The test would then be too difficult to deal with for our purposes, and it was for this reason that "?" was omitted from both forms.

Projection. From statements in the literature [1, 2] projection may be defined to include the following aspects: (*a*) it is defensive; (*b*) it is unconscious; (*c*) it involves undesirable characteristics which threaten the ego; (*d*) it occurs when the ego cannot repress the characteristics; and (*e*) it is a self-deception and therefore a reality distortion. An operational approach was therefore formulated to encompass this definition, as follows. The subject, in answering the first form (about himself) denies that he possesses a certain trait. On the second form, however, he says that others do possess this trait. But it is entirely possible that others *do* possess it, and that he is realistic about others and is not projecting. He *is* projecting if the majority (51% or more) of the group, in taking the test for themselves the first time, say that they do *not* have the trait in question. In other words, if the subject says, "I don't have this, but others do," when actually the majority of others say they do not, then projection is occurring. This procedure was done only for those questions which involved undesirable traits since by definition projection involves undesirable characteristics which the subject cannot repress and against which he unconsciously defends himself. He is being unrealistic in attributing the trait to others when they say they do not possess the trait. This approach is supported by Remmers [14] who, in defining projection, says that if operationally there is a significant difference between the actual measurement of the characteristic and the estimate given by the subject then projection is operative.

Each subject was given a point for projection each time he answered a question according to the above formula. Total scores on projection were obtained for only three of the GAMIN factors—A, I, and N. Trait A is ascendancy in social situations as opposed to submissiveness; I is lack of inferiority feelings or self-confidence; and N is lack of nervous tenseness and irritability. Guilford and Martin state that high scores on G, A, I, and N mean generally unfavorable characteristics and low scores mean favorable characteristics [11]. However, G (general pressure for activity) may possibly be equivocal for this study, since it measures either a tendency to engage in vigorous overt action or a disinclination for motor activity. Likewise, trait M (Masculinity) may be equivocal since a high degree of femininity is to be expected among college males [5]. Therefore, G and M were not used, since they are not as ego-involved as A, I, and N.

Empathy. Dymond [6, 7] defines this as an imaginative transposing of oneself into the thinking, feeling, and acting of another, and with Cottrell [4] she says it means to take another's role, place oneself in his shoes, and perceive the situation from his perspective. Grossman [10] refers to it as perceiving correctly the world from the other person's frame of reference. This perhaps involves projection in that one projects oneself into the thinking, feeling, and acting of another. The difference between this type of transposition and true projection lies in the realistic nature of empathy. Remmers [14], as noted above, has said that operationally if there is a significant difference between the actual measurement of the characteristic and the estimate given by the subject, then projection is occurring. However, he says, if the measurement of the characteristic and the estimate given by the subject are the same, then it is empathy. Thus it can be seen that empathy requires a realistic estimate of others whereas projection does not. The difference then between the projection which occurs in empathy and projection as defined for this study lies in the fact that in empathy the projection is one which is a true estimate. This reasoning is further justified by the fact that Sears [15] found that those who project lack insight, and Dymond [6] says that those who are empathic have insight, for to possess insight implies a realistic view of oneself and others.

Thus the operational definition for empathy used here is as follows. An individual empathizes when he says others possess a trait in question and others actually do have it when speaking of themselves. In other words, the subject in taking the second form of the test says that others possess a trait; if the majority (51% or more) of the group, in taking the test for themselves (first form), agree with him, then it may be said that empathy is occurring. One point for empathy was given each time the subject answered a question according to this formula. There is then, as Remmers [14] says, no difference between the estimate of the measurement of others and the actual measurement of others.

Reality. Realism or reality may be used in several ways, i.e., a realistic outlook may imply that the individual looks at himself realistically or that he looks at the world as others do. For this study, as did Norman [13], we chose the latter aspect, a comparison of the individual's estimate of others with the group's estimate of others. The term as used here has been called insight by others. Steinmetz [16], in an objective study of insight, has defined it as the ability to perceive the world as others do. Thus the term called reality in the present study could be called insight since there is a comparison of how the individual's perception of others conforms to the group's perception of others. However, it has been shown [1, 6, 7] that this is only one aspect of insight, for in addition to being able to look at the world as others do, insight means being able to look at oneself objectively. Thus it was decided that a more adequate term for the present measure was reality. Operationally in this study, then, reality occurs when the individual states others have the trait in question and others actually do have it when the group is speaking of others. Specifically, the subject in answering the second form (about others), says they possess a certain trait; the majority of the group (51% or more) when taking the second form, also say that others possess the trait. The subject is then being realistic, i.e., he is seeing others as others see others — there is a conformity between the subject's and the group's estimate of others. He is given a point for reality each time this occurs.

For both total empathy and reality scores, all the questions for factors A, I, and N were used, since it is possible to empathize and be realistic on both desirable and undesirable traits. Also it was possible to obtain a score for both empathy and reality on both "Yes" and "No" answers, which was not possible with projection.

In testing the first hypothesis, it was necessary to beware of the fact that spuriously high correlations could result from intratrait comparisons. As noted above, projection scores were derived partially from the first form of the test when the group as a whole took the test answering for themselves. Empathy scores were also partially derived from this measure. Therefore, high intratrait correlations might result because both projection and empathy scores were derived from the same base. For example, high relationships might occur when relating A empathy to A projection, I empathy with I projection, etc. Hence, intertrait relationships must be examined, correlating every variable with every other. Therefore, A projection was correlated with I projection, I reality, I empathy, N projection, etc. Then, if significant correlations occurred between any variable derived from one trait and that derived from another, a further check on the first hypothesis could be offered. In addition, although this was not stated as a hypothesis, it could be empirically determined whether, if an individual projected in one trait, he would project in another, or if he were realistic in one trait, he would be realistic in another, etc.

To examine the first hypothesis, then, scores in projection, empathy, and reality were intercorrelated for both intertrait and intratrait comparisons. To test the second hypothesis, that reality and empathy would be positively correlated with adjustment, correlations were derived between scores in reality and empathy for traits A, I, and N, and actual test scores[1] obtained for these traits. The actual scores were the operational measures of adjustment. The hypothesis was first entertained that projection would correlate negatively with adjustment. This was not entirely tenable because it is obvious that in scoring a test like GAMIN, when a subject denies undesirable traits in himself, he obtains a better "adjustment" score. This is correct, because all intratrait correlations, as shown later in Table 3, have a positive *r* between projection and adjustment, a result which may be an artifact of the scoring of the test.

---

[1] These scores, of course, do not represent completely the scores the subject would have made had he been allowed the "?" alternative in answering the questionnaire the first time.

Results

Tables 1 and 2 present correlations testing the first hypothesis. Table 1 shows that all intratrait correlations are significant at the 1% level in the predicted direction. The median of the three intratrait correlations (excluding total scores) between projection and reality is —.45; the median of the 3 *r*'s between projection and empathy is —.65; and the median of the 3 *r*'s between reality and empathy is .83. Using total scores, these medians are respectively —.41, —.65, and .77. Regarding intratrait correlations, then, it may be seen that the first hypothesis is readily supported. However, as noted above, these *r*'s may have an element of spuriousness to them. Table 2, which shows intertrait correlations, presents the following data. There are six correlations between projection and reality, with a median of —.16, which is not significant. Only one of these six correlations is statistically significant, but all are in the direction predicted. Between projection and empathy, there are also six correlations, with a median of —.29, significant at the 5% level. Five of these six correlations are statistically significant, and all six are in the direction predicted. Between reality and empathy, there are again six correlations, the

Table 1
Intratrait Correlations among Experimental Variables

Relationship	*r*
A projection vs. A reality	—.35**
A projection vs. A empathy	—.41**
A reality vs. A empathy	.74**
I projection vs. I reality	—.45**
I projection vs. I empathy	—.65**
I reality vs. I empathy	.83**
N projection vs. N reality	—.49**
N projection vs. N empathy	—.66**
N reality vs. N empathy	.89**
Total projection (A + I + N) vs. total reality (A + I + N)	—.41**
Total projection (A + I + N) vs. total empathy (A + I + N)	—.65**
Total reality (A + I + N) vs. total empathy (A + I + N)	.77**

** Significant at the 1% level.

Table 2
Intertrait Correlations among Experimental Variables

Relationship	*r*
A projection vs.	
I projection	.50**
I reality	—.03
I empathy	—.23*
N projection	.41**
N reality	—.17
N empathy	—.22
A reality vs.	
I projection	—.17
I reality	.62**
I empathy	.30**
N projection	—.14
N reality	.26*
N empathy	.25*
A empathy vs.	
I projection	—.35**
I reality	.30**
I empathy	.50**
N projection	—.23*
N reality	.16
N empathy	.26*
I projection vs.	
N projection	.57**
N reality	—.29*
N empathy	—.37**
I reality vs.	
N projection	—.13
N reality	.48**
N empathy	.37**
I empathy vs.	
N projection	—.49**
N reality	.39**
N empathy	.52**

* Significant at the 5% level.
** Significant at the 1% level.

median of which is .30, significant at the 1% level. Again, five of these six *r*'s are statistically significant, and all are in the direction predicted by the hypothesis. Of the total of 18 intertrait *r*'s, all or 100% are in the indicated direction; 11 or 61% are statistically significant either at the 1% or 5% level, a figure which far exceeds chance expectations. The first hypothesis is thus supported by data from both Tables 1 and 2.

Each of the variables also correlates well with itself, according to Table 2. The median of the three correlations of projection in one trait with projection in another is .50, significant at the 1% level. The median of the three *r*'s of reality with reality is .48; of the three *r*'s of empathy with empathy is .50. Both these *r*'s are significant also at the 1% level. Thus it may be stated that if an individual projects in one GAMIN trait, he will project in another, if he is realistic in one he will be realistic in another, etc.

Table 3

Correlations between Adjustment and Experimental Variables

Relationship	*r*
A adjustment vs.	
A projection	.43**
A reality	.09
A empathy	.27*
I projection	.11
I reality	.24*
I empathy	.29*
N projection	.00
N reality	.00
N empathy	.06
I adjustment vs.	
A projection	.24*
A reality	.25*
A empathy	.33**
I projection	.13
I reality	.44**
I empathy	.56**
N projection	—.01
N reality	.22
N empathy	.31**
N adjustment vs.	
A projection	.11
A reality	.33**
A empathy	.42**
I projection	—.02
I reality	.38**
I empathy	.43**
N projection	.13
N reality	.21
N empathy	.38**
Total adjustment (AIN) vs.	
Total projection (AIN)	.18
Total reality (AIN)	.46**
Total empathy (AIN)	.56**

* Significant at the 5% level.
** Significant at the 1% level.

The second hypothesis stated that reality and empathy will correlate more closely with adjustment than will projection. Data bearing on this hypothesis are presented in Table 3. In this table, the median intratrait *r* of the three correlations between projection and adjustment is .13 (that all three are positive may be an artifact of scoring the GAMIN, as stated before). The median of the three intratrait *r*'s between reality and adjustment is .21; the median of the three between empathy and adjustment is .38. Only the latter median figure is statistically significant, but the order is in the direction of our hypothesis. Projection and adjustment give one rather high statistically significant *r* of .43, for trait A. Since all of our subjects were males, it is perhaps expected that denial of lack of ascendance would be important. Reality also has one statistically significant intratrait *r* with adjustment, but empathy correlates significantly in all three instances. Using total scores, the *r* of total adjustment on traits AIN with total projection on these traits is .18, with total reality it is .46, with total empathy it is .56. The latter two are significant at the 1% level.

If we examine the intertrait *r*'s which are free of contamination, it will be noted that the median *r* between projection and adjustment is just about .05 for the six correlations. Only one is statistically significant. The median *r* for the relationship between reality and adjustment is .25, significant at the 5% level. Among these six correlations, four are statistically significant. The median of the six *r*'s between empathy and adjustment is .32, significant at the 1% level. Five of these six *r*'s are statistically significant.

Thus, for both intra- and intertrait relationships, there appears a hierarchy of order of correlation with adjustment as follows: projection lowest, reality next, and empathy highest. From the standpoint only of significant *r*'s with adjustment, projection has two, reality five, and empathy eight, according to Table 3. The second hypothesis, then, appears to be supported.

Discussion

Our hypotheses are generally upheld by the data, the method used in the present study giving valid results supporting current personality theory. Certain methodological objections may be raised, especially that of a lack of knowledge about the reliability coefficient of the judgments of others (i.e., when the test was taken the second time), the fact that adjustment scores were changed in some unknown quantity by the elimination of the "?" alternative, the comparatively low *N* which made a "51% criterion" only 38 cases, and the assumption that self-reports were valid measures of adjustment. The latter criticism has been particularly mentioned by Ellis [9]. However, despite these objections, it is the writers' belief that this operational approach to the investigation of personality dynamics can be quite fruitful. It is certainly no accident that despite some of the crudities of the present investigation, *all* the correlations in Tables 1 and 2 were in the expected direction, and the trend of the correlations in Table 3, whether intra- or intertrait, was in the nature hypothesized at the outset. Further research might well employ outside criteria of adjustment, projection, empathy, and reality, to test whether our operational definitions have any validity.

Summary

An attempt was made, using operational definitions of terms, to verify two hypotheses. The first was that projection would be negatively related to both reality and empathy, but that the latter two variables would be positively related to each other. The second was that reality and empathy would correlate more closely with adjustment than would projection. Both these hypotheses were generally verified, using a personality questionnaire, the GAMIN, given under conditions wherein the subject took it for himself first, then made judgments about others. Consistent intra- and intertrait correlations in the direction predicted supported the first hypothesis. The second was supported by correlations giving a hierarchy of relation to adjustment from greatest to least of empathy, reality, and projection.

Received May 1, 1953.

References

1. Alexander, F. *Fundamentals of psychoanalysis.* New York: Norton, 1948.
2. Allport, G. W. *Personality, a psychological interpretation.* New York: Holt, 1938
3. Bellak, L. The concept of projection, an experimental investigation and study of the concept. *Psychiatry,* 1944, **7**, 353-370.
4. Cottrell, L. S., & Dymond, Rosalind F. The empathic responses. *Psychiatry,* 1949, **12**, 355-359.
5. DeCillis, Olga E., & Orbison, W. D. A comparison of the Terman-Miles M-F Test and the *Mf* scale of the MMPI. *J. appl. Psychol.,* 1950, **34**, 338-342.
6. Dymond, Rosalind F. The relation of insight and empathy. *J. consult. Psychol.,* 1948, **12**, 228-233.
7. Dymond, Rosalind F. A scale for the measurement of empathic ability. *J. consult. Psychol.,* 1949, **13**, 127-133.
8. Dymond, Rosalind F. Personality and empathy. *J. consult. Psychol.,* 1950, **14**, 343-350.
9. Ellis, A. The validity of personality questionnaires. *Psychol. Bull.,* 1946, **43**, 385-440.
10. Grossman, D. The construction and validation of two insight inventories. *J. consult. Psychol.,* 1951, **15**, 109-114.
11. Guilford, J. P., & Martin, H. G. *The Guilford-Martin inventory of factors GAMIN.* Beverly Hills, Calif.: Sheridan Supply Co., 1943.
12. Healy, W., Bronner, Augusta F., & Bowers, A. M. *The structure and meaning of psychoanalysis.* New York: Knopf, 1945.
13. Norman, R. D. The interrelationships among acceptance–rejection, self–other identity, insight into self, and realistic perception of others. *J. soc. Psychol.,* 1953, **37**, 205-235.
14. Remmers, H. H. A quantitative index of social-psychological empathy. *Amer. J. Orthopsychiat.,* 1950, **20**, 161-165.
15. Sears, R. R. Experimental studies of projection: I. Attribution of traits. *J. soc. Psychol.,* 1936, **7**, 151-163.
16. Steinmetz, H. C. Directive psychotherapy: measuring psychological understanding. *J. clin. Psychol.,* 1945, **1**, 331-335.

Journal of Consulting Psychology
Vol. 18, No. 1, 1954

Moral Indignation: A Correlate of Denied Dislike of Parents' Traits

Sidney M. Jourard
Emory University

The hypothesis under consideration in the present study is stated as follows: *S*s who will admit to a predominant degree only favorable attitudes toward many traits of their parents will express a higher degree of moral indignation than *S*s who will express some attitudes of dislike toward their parents' traits.

Moral indignation is defined as the expression of annoyance, anger, or hostility toward transgressors of social mores.

The rationale for the hypothesis may be stated as repression-displacement theory of hostility. According to this theory, hostility toward themselves is inevitably generated by parents in their efforts to socialize their children. This hostility in the children may be allowed or denied expression. If it is denied, the children will learn to repress it, and displace it to hate objects provided by a culture for the guilt-free expression of hostility. Further, where hostility toward parents has been repressed, the children will tend to perceive the parents in undifferentiated, uncritical, almost magical terms, as Frenkel-Brunswik [1, pp. 340-346] found to be the case in highly authoritarian subjects whom she interviewed.

Procedure

Materials. Three questionnaires were devised to measure the relevant attitudes, viz: a moral indignation scale (MI), a scale to measure the feelings of each *S* toward 61 traits of his father (the Father-cathexis scale, C_F), and an identical scale to measure feelings toward the mother (Mother-cathexis scale, C_M).

Moral indignation scale.[1] Fifty statements were selected which seemed to describe "breaches of common decency" and which would stimulate moral indignation in a person. Many of these were paraphrased from items used in the F scale [1] and from Cason's study of annoyances [2], while others were devised by the writer.[2] The *S*s were required to indicate on the following five-point scale the degree of annoyance provoked by each situation:

1. Not annoyed
2. Slight irritation
3. Annoying
4. More annoying
5. Makes me furious

Sample items are as follows: A communist making a speech to you about his beliefs. A person who avoids the army draft because of "nervous" illness. The rapidity with which Hollywood stars change spouses.

Father-cathexis and Mother-cathexis scales.[2] Sixty-one traits were listed, and *S* was required to signify the nature of his feelings toward each of these traits, as he knew them in his father and mother.

The scale was devised as follows:

1. Have strong positive feelings: encircle a 1 for those aspects of the person which you like very much.

---

[1] The questionnaire forms have been deposited with the American Documentation Institute. Order Document No. 4032 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting $1.25 for 35 mm. microfilm or $1.25 for 6 x 8 in. photocopies. Make checks or money orders payable to: Chief, Photoduplication Service, Library of Congress.

[2] Many ideas for items were obtained by perusal of letters to the editor of the local newspaper where manners and morals are commented upon, often with moral indignation. Finally, one important source of ideas for items deserves mention — the topics about which the writer, his colleagues, and many housewives gossip when gathered for social purposes.

2. Moderate positive feelings: encircle a 2 for those traits which you like in the person, but not as strongly as in category 1.

3. Have no feelings one way or the other: encircle a 3 for those aspects of the person about which you have no feeling.

4. Have moderate negative feelings: encircle a 4 for those aspects of the person which you dislike moderately.

5. Have strong negative feelings: encircle a 5 for those aspects of the person which you strongly dislike.

Sample items from the questionnaires are: sense of humor, temper, general knowledge, ability to express sympathy.

Subjects. One hundred and fifteen *S*s served in the study. These included 69 male college undergraduates, 24 college females, 17 females from a YWCA social club, and 5 Air Force men tested by a student assisting the writer. The age range was from 18 to 42, with a mean of 23 and an *SD* of 4.23.

Testing procedure. All the questionnaires were administered at one sitting—the MI first, and then the C_F and C_M scales. Instructions were as follows:

> On these scales, you are asked to indicate your feelings. Please fill them out with as much self-honesty as you can. To assist you in this, we are asking you only to put down your birth date, so you need not fear what anyone might think concerning how you respond. Remember, please, to examine your feelings, so that your answers reflect these as accurately as possible.

Results

Means, standard deviations, range of scores, and odd-even reliability coefficients for the three tests are shown in Table 1.

All three tests appear to have high reliability and a broad range of scores.

The general hypothesis may be restated as follows: there will be a negative correlation between MI scores and both C_F and C_M scores.

The moral indignation scores were found to correlate −.39 with Father-cathexis, and −.37 with Mother-cathexis. Both *r*'s, while low, are significant beyond the .001 level, and support the hypothesis.

Discussion

The present data signify that persons who admit predominantly favorable attitudes to their parents tend to express a high degree of moral indignation, as defined and measured herein. Persons able to admit more critical feelings (higher C_F and C_M scores) toward their parents tend to produce lower MI scores.

Table 1

Means, Standard Deviations, Ranges, and Reliability Coefficients for the Three Questionnaires

	Mean	σ	Range	Odd-even reliability coefficients*
Moral indignation	136.36	22.90	86–192	.90
Mother–cathexis	137.48	29.50	80–209	.95
Father–cathexis	136.38	28.08	67–206	.95

* Correlations are between odd and even halves of tests and are corrected by the Spearman-Brown formula.

It must be noted, however, that the correlations, despite their high confidence level, are low; a very small amount of the variance in MI scores is accounted for by variation in the parent-cathexis scores. This should not be surprising since there might be many reasons for the expression of moral indignation besides the displacement of repressed resentment of parents.

Summary

One hundred and fifteen persons were tested with scales designed to measure moral indignation and attitudes toward parents. It was hypothesized that *S*s who admit little criticism and dislike of their parents would show higher indignation scores than *S*s who do admit some dislike of their parents' traits. Low but highly significant correlations between MI scores and C_M and C_F scores supported this hypothesis.

Received May 18, 1953.

References

1. Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. Cason, H. Common annoyances, a psychological study of every-day aversions and irritations. *Psychol. Monogr.,* 1930, 40, No. 2 (Whole No. 182).

Journal of Consulting Psychology
Vol. 18, No. 1, 1954

The Comparability of Measures of Profile Similarity: An Empirical Study

James N. Mosel
The George Washington University

and June B. Roberts
Pennsylvania State College

In recent years there has been an increased emphasis on the use of psychometric profiles in clinical diagnosis, vocational guidance, personnel evaluation, and even social psychology. This increased emphasis has resulted largely from the realization that interpretation of an individual's set of scores, for diagnosis or prediction, must frequently be based upon the pattern or configuration of scores, rather than upon consideration of one test at a time or upon an additive sum of scores. Computation of mean score differences between nosological or criterion groups, or the derivation of linear combinations of test scores to predict a criterion (as in the multiple-regression equation and discriminant function), exemplify the latter, nonconfigural approach. It has been felt by many, especially clinicians, that these conventional techniques are inadequate because they fail to take into account the fact that the diagnostic or predictive meaning (i.e., weight) of a score on one test is not constant, but varies according to the scores obtained on the other tests. Consequently, the argument goes, interpretation must consider patterns of scores. This is mathematically tantamount to saying that interpretation involves a nonlinear combination of scores—nonlinear in that the weights assigned to each test are not constant but vary as a function of score on the test itself and on the other tests. It is this set of intertest relationships as defined by a nonlinear combination that is implied by such terms as "pattern," "configuration," "profile effect," etc. Discussion of the problems and concepts of configural vs. linear treatments of multiple measurements will be found in Horst [6] and Mosier [9].

The great interest in patterns has quite naturally led to the problem of matching profiles—measuring the degree of similarity in the pattern of two profiles. The usual situation in clinical practice in which this need arises is where a standard or reference profile has been established[1] and it is desired to match a given "unknown" or referred profile against it. The assumption underlying this is that the closeness of the match measures the extent to which the unknown profile will display certain criterion or nosological characteristics associated with the standard profile.

The need for profile matching also arises in the research situation when it is desired to learn whether the profiles of one diagnostic category are more similar to each other than they are to the profiles of some other diagnostic category. The principle here is that if the within-groups similarity is not greater than between-groups similarity, the groups cannot be distinguished by profile patterns. This is illustrated in a study by Kogan [7], who attempted to discriminate between schizophrenic and neurotic patients by means of Wechsler subtest patterns. In a similar vein, *Q* technique attempts to determine whether the agreements among the profiles of persons are such that when factor analyzed, a typology can be established. Investigations of this type, incidentally, provide one approach to the clinical problem of establishing the standard or reference profile mentioned above. Before a standard profile can be established, it must first be shown that a basic profile exists which is typical of the criterion group and which differentiates this group from others.

---

[1] The problem of identifying and specifying standard profiles is beyond the purview of this paper, but it is quite apparent that systematic methods are yet to be developed. It should be noted that the conventional procedure of computing the mean scores of a given criterion group and then using the resulting profile of averages as the standard completely ignores the possibility that the number of characteristic profiles may be greater than one.

There are two aspects of profiles on which matching may occur: the *shape* or configuration of scores and the general *elevation* (also called intensity or level) of scores as indicated by some measure of central tendency. Thus we may speak of matching for *absolute* agreement, in which we consider similarity of both shape and elevation, or for *relative* agreement, in which we consider only similarity of shape.

Considerable discussion in the literature [2, 3] has been devoted to the relative merits of measuring similarity in terms of shape and elevation or of shape alone. The basic question is whether profiles of similar shape but of different elevations indicate the same diagnosis or prediction. Much of the confusion apparently stems from the failure to realize that there is no single, best conception of profile similarity and that the decision between the two alternatives rests on what particular characteristics interest the investigator and on what empirical studies show he can obtain. Profiles can yield several types of information, some of which are perhaps more clearly related to the shape of the profile, while others are more closely associated with both shape and elevation. In many instances the ultimate answer can be given only by empirical research designed to show which approach provides the greatest differentiation with respect to an external criterion reflecting the investigator's purpose.

One interesting variation of the empirical approach is that of Barnette [1], who evaluated separately differences in shape and elevation of aptitude profiles. That this approach may be fruitful is further supported by Schmidt's work [14] with the MMPI. He found that groups diagnosed as severe and mild psychoneurosis had highly similar profile shapes but different elevations, the severe group having a more elevated profile than did the mild.

There are, of course, certain situations where no outside criterion is available or even meaningful, such as in Osgood's use of profiles to display connotative meaning [10]. In such cases an entirely rational decision is not only necessary but proper.

Indices of Profile Similarity

In an attempt to overcome the obvious disadvantages of matching profiles by visual judgment, a number of statistical indices of profile similarity have been proposed. These are described below, but because the present study is concerned with the empirical agreement among these, no attempt will be made to elaborate upon or evaluate their statistical properties.[2]

Correlation coefficient. Perhaps the most obvious and generally employed statistic is the product-moment or rank-difference correlation coefficient. The correlation technique yields a nonarbitrary measure which reflects similarity of shape but disregards elevation.[3] This index ranges from 1.00 for high similarity, through zero for no similarity, to − 1.00 for high inverse similarity (i.e, when the two profiles are mirror images of one another). The correlation coefficient has a disadvantage in that profiles of rather dissimilar shape can have equal correlation with the standard profile. Also, the value of 1.00 does not necessarily indicate perfect similarity; two profiles can be quite dissimilar in shape and still correlate perfectly.

Meehl's difference index. This measure [8] reflects shape but neglects elevation. The computational procedure is as follows:

1. Obtain for each profile the deviations (keeping algebraic sign) of the scores of each test from the second highest test in the profile. These values may be symbolized as d.
2. Obtain the differences between the corresponding deviations in the two profiles. These values may be symbolized as δ.
3. Square the δ's after first rounding to the nearest tenth.[4]
4. Sum the squared δ's of the N tests and extract the square root.

In formula form:

$$I = \sqrt{\sum_{i}^{N} \delta^2_{jl}}$$

where $\delta_{jl} = d_{ij} - d_{il}$; and $d_{ij} = X_{ij} - S_j$ and $d_{il} = X_{il} - S_l$, in which X_{ij} and X_{il} are corresponding scores of profiles j and l on test i (of N tests), and S_j and S_l are the second highest scores in profiles j and l respectively.

Meehl states that his index tells how far apart two profiles are in the hyperspace of N dimensions in which each profile is projected as a point. The smaller the value the closer the match, although the index is an arbitrary one, having no absolute meaning. Since its value depends on the number of tests, it is comparable only for matchings in which all pairs of profiles have the same number of tests. Meehl reports that the index correlates .68, .65, and .91 with subjective estimates of similarity made visually by three judges. This index is very new and Meehl admits that it is frankly experimental and that its statistical properties remain to be determined.

---

[2] For a general discussion of various indices, see Cattell [2], Du Mas [3], and Gaier and Lee [5].

[3] Intraclass correlation, however, considers both shape and elevation differences [16]. This is because the deviations are taken from the grand mean of both sets of profile scores. On the other hand, this index suffers from a restricted negative range.

[4] Meehl [8] reports that the rounded values correlate .95 with the unrounded.

D coefficient. Osgood and Suci [11] have described an index designed to reflect both shape and elevation. Although these authors apparently consider their proposal as new, the same method was suggested some time ago by Sagebeer [13] and more recently by Toops [15], who calls it the "precise profile method." The D coefficient is based on the geometric principle that in a space of N mutually orthogonal dimensions, the distance between two points is equal to the square root of the sum of the squared differences between the coordinates of the points on each dimension. Since profiles may be considered as points in N space (where N equals the number of tests), the distance between them serves as a measure of similarity. This distance is D, which in formula form is:

$$D = \sqrt{\sum_{i}^{N} d_{jl}^{2}}$$

where $d_{jl} = X_{ij} - X_{il}$, in which X_{ij} and X_{il} are the corresponding scores of profiles j and l on test i of N tests. The computational procedure is simple:

1. Obtain the difference between corresponding scores on each test for the two profiles.
2. Square and sum these differences.
3. Extract the square root.

Like Meehl's index, the D coefficient is an arbitrary number whose value depends on the number of tests in the profile. It should also be noted that the above formula applies only when the tests are uncorrelated.

Du Mas' coefficient of profile similarity. Du Mas [4] has introduced the coefficient of profile similarity, r_{ps}, which describes similarity in terms of the agreement in slope between corresponding segments of the two profiles being matched. By "segment" is meant the line drawn from one score point on the psychograph to another adjacent point. The slope is positive when the left score point of a segment is closer to the bottom horizontal of the graph than is the right score point. Zero slope occurs when the segment is parallel to the bottom horizontal.

In formula form:

$$r_{ps} = 2\ (S/T - .50),$$

where S is the sum of the corresponding profile segments which exhibit slope of the same sign, and T is the total number of segments in which no zero slope occurs. The standard error of r_{ps} is given by Du Mas as

$$\sigma_{r_{ps}} = \sqrt{1/T},$$

thus enabling a test of the null hypothesis. Du Mas [4] states that the assumptions underlying the coefficient are that the test score distributions are symmetric and identical, the regression of one test on another adjacent test is linear, and the sequence of the tests remains fixed.

Obviously, r_{ps} considers only profile shape and, in addition, loses precision because it is based only upon the sign of the slope, not the magnitude. Consequently, it is quite possible, as Du Mas himself points out, to obtain a very high index of similarity when the two profiles are quite dissimilar in both shape and elevation. While r_{ps} can range from -1.00 to 1.00, it cannot, however, always assume a zero value. Moreover, it yields a discrete rather than continuous distribution of values. The number of possible values it can assume can never be greater than $N + 1$ (where N is the number of tests in the profile). For a four-test profile r_{ps} can assume only five values: 1.00, 0.50, 0.00, $-$ 0.50, and $-$ 1.00. Also, the fact that the index is an arbitrary function of the order in which the tests are arranged may constitute a criticism. This feature is not shared by the other indices described here. The index has the advantage, however, of being rapid to compute, especially in view of the table Du Mas gives for cases where $T \leqslant 30$.

Cattell's coefficient of pattern similarity. This coefficient, r_p, is one of three indices developed by Cattell [2] to measure different aspects of profile similarity, the other two being his coefficient of pattern shape, r_s, and his coefficient of pattern-effect, r'_p. He concluded that the most generally applicable is r_p since this index is designed to reveal both profile shape and elevation.

The present paper will therefore be confined to r_p which is defined by the formula

$$r_p = \frac{2k - \Sigma d^2}{2k + \Sigma d^2}$$

where k is the median for chi square (i.e., the value of chi square at $p = .50$) with degrees of freedom equal to $N - 1$, where N is the number of tests in the profile; and d is the difference between corresponding scores on the two profiles at each test, expressed in standard score form with respect to the distribution within that test.

Cattell states that for practical purposes the index can be considered as approximately similar to an ordinary correlation coefficient. It ranges from 1.00, meaning perfect similarity, through zero, meaning complete lack of relation, to −1.00, meaning complete inverse relationship. The assumptions are that the tests are independent and that the distribution of each test is normal.

Purpose of the Present Investigation

In view of the above multiplicity of measures, a crucial question relates to the amount of agreement these measures would display in rank ordering a set of profiles in similarity to a standard profile. The purpose of the present study was to determine empirically the comparability of these measures as revealed by their intercorrelations. Inasmuch as these measures were originally devised to replace the presumably less reliable method of subjective matching, an additional purpose was to compare the statistical indices with the results from visual judgment. The question also arises as to whether the agreement among indices varies according to the standards used. Correlations were therefore computed separately for matchings in which differing profiles were used as standards.

Procedure

Actual profiles were used, consisting of scores on the first five subscales of the Kuder Preference Record as obtained from a group of supervisors in a government agency. Four profiles were selected as standards because of their extreme visual dissimilarity. A group of 30 other profiles served as "referred" profiles. Using the five indices described above, measures of similarity to each standard were computed for each referred profile.

Subjective Matching

In measuring similarity by visual judgment, an effort was made to secure reliable judgments. Apparently no studies have been made of the reliability of judges or of methods in rating the similarity of profiles; consequently, several preliminary studies were made using two judges and three rating methods. Judge A was a psychologist; judge B, an engineer. For whatever significance it may have, both judges were above the 85th percentile on the Minnesota Paper Formboard.

Method of single stimuli. The judges rated each profile against each standard on the following six-point scale: 1: "almost identical," 2: "very similar," 3: "fairly similar," 4: "somewhat similar," 5: "no resemblance," 6: "completely opposite." To permit the judge to establish a stable subjective scale of judgment, all profiles were matched against one standard before rating against the next standard was begun. Each judge repeated his ratings one week later.

Rank-order method. Next, each profile was ranked in order of similarity to each standard. The judges first rank ordered the profiles previously assigned to each rating-scale interval. When profiles had thus been ranked within each group, they could readily be placed in an over-all rank order of similarity to the standard.

Method of triads. As a further check on the reliability of the above rank orders, judge A assigned a second set of rank-order values using the method of triads [12]. In this method the judge compares *two* referred profiles with the standard and decides which of the two most resembles the standard. As in the method of paired comparisons, all possible pairs of referred profiles are thus compared with the standard. The total number of times each profile is judged most similar can then be used to yield rank orders.

Results

Judgmental matching. Table 1 shows the rate-rerate reliabilities of the two judges from the method of single stimuli (correlations expressed as Pearson r) and from the method of rank order (correlations expressed as rho). Table 2 shows the interjudge agreement for the two methods, basing the intercorrelations on the mean of each judge's first and second ratings.

Table 1
Rate-Rerate Reliabilities of Two Judges in Matching 30 Profiles against Four Standard Profiles by the Method of Single Stimuli and the Method of Rank Order

Standard	Judge A		Judge B	
	Single stimuli*	Rank order	Single stimuli	Rank order
1	.75	.87	.85	.87
2	.91	.95	.88	.90
3	.92	.95	.70	.78
4	.83	.76	.46	.25

* The correlations for the method of single stimuli are given in Pearson r; those for the method of rank order in rho.

Table 2
Intercorrelation between Two Judges in Matching 30 Profiles against Four Standards by the Method of Single Stimuli and the Method of Rank Order

Standard	Single stimuli (r)	Rank order (ρ)
1	.28	.53
2	.80	.85
3	.50	.67
4	.71	.58

It will be noted that the rank-order method tends to yield higher reliabilities than does the method of single stimuli. As might be expected, the interjudge reliabilities are consistently lower than the rate-rerate reliabilities.

These results indicate that both judges are fairly self-consistent but differ somewhat in their definitions of "similarity." Judge A has a very acceptable reliability, being somewhat superior to judge B. Judge A's reliability is further evidenced by a rho of .80 between the mean of his first and second values from the rank-order method and the method of triads. This correlation is almost as great as A's rate-rerate reliability by the rank-order method. The conclusion is that the ratings of judge A by the rank-order method are the most reliable. These were therefore retained for comparison with the various statistical indices of profile similarity.

Index agreement. Table 3 shows the matrix of intercorrelations (rank difference) among the five statistical indices of profile similarity and the subjective judgments of judge A. The correlations have been determined separately for each standard profile.

Table 3
Intercorrelations among Five Statistical Measures and Visual Judgments of Profile Similarity; 30 Profiles Matched against Four Standard Profiles

(Correlations expressed as rho)

Measure	Meehl	Osgood	Du Mas	Cattell	Pearson
		Standard 1			
Osgood	.73				
Du Mas	.005	.30			
Cattell	.81	.51	.04		
Pearson	.70	.28	.17	.60	
Judge A	.65	.41	.50	.78	.49
		Standard 2			
Osgood	.81				
Du Mas	.82	.73			
Cattell	.65	.64	.73		
Pearson	.80	.71	.74	.54	
Judge A	.80	.77	.90	.68	.75
		Standard 3			
Osgood	.60				
Du Mas	.49	.46			
Cattell	.32	.31	.48		
Pearson	.14	.42	.20	–.006	
Judge A	.29	.28	.57	.55	–.002
		Standard 4			
Osgood	.65				
Du Mas	.25	–.11			
Cattell	.59	.48	.12		
Pearson	.54	.40	.23	.50	
Judge A	.38	.06	.40	.47	.17

It is obvious that the type of standard employed makes a sizable difference in the amount of agreement among the measures. For instance, the Du Mas and Meehl indices correlate .005 when the profiles are matched against Standard 1, but .82 when Standard 2 is used. This means, of course, that the extent to which these measures are tapping a common factor of "similarity" varies according to the profile adopted as the standard. The comparability of the measures cannot, therefore, be stated in absolute terms; it depends on the standard against which the profiles are matched.

Inspection of Table 3 reveals that the two measures which disagree most with the other measures are the Du Mas coefficient[5] and Pearson r, especially the latter. Cattell's coefficient agrees most closely with subjective judgments of similarity, while the indices of Du Mas, Meehl, Osgood, and Pearson follow in that order. The indices of Meehl and Osgood intercorrelate more highly on the whole than do any other pairs of measures. This is somewhat surprising in view of the fact that the former reflects only profile shape, while the latter reflects both shape and elevation.

Summary

The agreement among five statistical indices of profile similarity and visual judgments was studied through intercorrelation. A group of 30 profiles from the Kuder Preference Record was matched against four different standard profiles, and the resemblance was measured by Cattell's coefficient of pattern similarity, Du Mas' coefficient of profile similarity, Osgood and Suci's D coefficient, Meehl's index, and Pearson r. Visual judgments of profile similarity by two judges using two methods (single stimuli and rank order) showed acceptable rate-rerate reliabilities; the interjudge reliabilities were somewhat lower. The scale values from the most reliable judge using the most reliable method (rank order) were adopted for comparison with the results from the five statistical indices.

Results showed a wide degree of disagreement among the measures as to the rank order in which they arranged a set of profiles with respect to similarity to the standard profile, the rank-difference intercorrelations ranging from −.006 to .90. Also, the amount of agreement between any two measures of profile similarity varied considerably according to the particular profile employed as the standard, thus indicating that the comparability of the measures cannot be stated in absolute terms but is contingent upon the standard used.

Cattell's coefficient showed the closest agreement with visual judgments, while the indices of Du Mas, Meehl, Osgood and Suci, and Pearson followed in that order. The greatest agreement among the statistical indices was that between the coefficients of Meehl and Osgood.

Received April 1, 1953.

References

1. Barnette, W. L. Occupational aptitude patterns of selected groups of counseled veterans. *Psychol. Monogr.*, 1951, **65**, No. 5 (Whole No. 322).
2. Cattell, R. B. r_p and other coefficients of pattern similarity. *Psychometrika*, 1949, **14**, 279-298.
3. Du Mas, F. M. On the interpretation of personality profiles. *J. clin. Psychol.*, 1947, **3**, 57-65.
4. Du Mas, F. M. The ciefficient of profile similarity. *J. clin. Psychol.*, 1949, **5**, 123-131.
5. Gaier, E. L., & Lee, Marilyn. Pattern analysis: the configurational approach to predictive measurement. *Psychol. Bull.*, 1953, **50**, 140-148.
6. Horst, P. The prediction of personal adjustment. New York: *Soc. Sci. Res. Coun. Bull.*, 1941, No. 48.
7. Kogan, W. S. An investigation into the relationship between psychometric patterns and psychiatric diagnosis. *J. gen. Psychol.*, 1950, **43**, 17-46.
8. Meehl, P. E. *Minnesota Multiphasic Personality Inventory research for counselors.* St. Paul: St. Paul Dept. of Educ., 1950.
9. Mosier, C. I. Batteries and profiles. In E. F. Lindquist (Ed.), *Educational measurement.* Washington, D. C.: Amer. Coun. on Educ., 1951. Pp. 764-808.
10. Osgood, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 197-237.
11. Osgood, C. E., & Suci, G. J. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, **49**, 251-262.
12. Richardson, M. W. Multi-dimensional psychophysics. *Psychol. Bull.*, 1938, **35**, 659. (Abstract)
13. Sagebeer, R. Identifying patterns of ability. *Educ. Rec. Suppl.*, 1938, **19**, 172-178.
14. Schmidt, H. O. Test profiles as a diagnostic aid: the MMPI. *J. appl. Psychol.*, 1945, **29**, 115-131.
15. Toops, H. A. Philosophy and practice of personnel selection. *Educ. psychol. Measmt*, 1945, **5**, 95-124.
16. Webster, H. A note on profile similarity. *Psychol. Bull.*, 1952, **49**, 538-539.

---

[5] The correlation of Du Mas' r_{ps} with the other indices is undoubtedly reduced because of the restricted range of values which this index can assume. No correction for widespread classes was made, however, because it was desired to ascertain the interrelationships in terms of the categories given.

Journal of Consulting Psychology
Vol. 18, No. 1, 1954

Children's Responses to Animal and Human Figures in Thematic Pictures[1]

Mary Ann Smith Armstrong
Institute of Child Welfare, University of Minnesota[2]

The Children's Apperception Test (CAT) [1] is a projective test for children primarily between the ages of three to ten years. The CAT is based on the theory that children can identify more readily with animal figures than with human figures.

The problem of this study was to determine which picture series elicits more thematic material. The hypotheses of the writer are as follows:

1. Pictures which contain human figures will evoke longer fantasy productions.
2. More verbalization which is nondescriptive in content will occur in protocols based on human figures.
3. Nouns, verbs, and ego-related words will have higher frequencies in stories having human figures as their main characters.
4. Responses to pictures depicting human relationships will involve more "blocking," which, in turn, will produce a longer reaction time.

Method

The subjects were 60 children who attended the first, second, and third grades in the University of Minnesota elementary school. The sample included 10 boys and 10 girls from each grade. The mean age of the first grade subjects was 6-7, with a range from 6-2 to 7-3. Second grade subjects ranged in age from 7-3 to 8-2, with a mean age of 7-6. Children in third grade had a mean age of 8-6, with a range from 8-1 to 8-11. The mean IQ of first grade subjects measured by the 1937 Revision of the Stanford-Binet was 120, with a range from 76 to 156. Children in the second grade had a mean IQ of 131, with a range from 108 to 174. Third grade subjects ranged from 97 to 172, with a mean IQ of 123. The subjects were predominately of upper socioeconomic status.

Bellak's CAT pictures were ranked by 10 judges on the basis of their supposed stimulus value in evoking verbalizations from children. Graduate students and faculty served as raters. Five pictures from the original CAT were selected for use in this study on the basis of the mean ratings obtained. The experimental pictures (Series B) depicted a feeding situation (Card 1), attitudes toward parents (Card 2), sibling rivalry (Card 4), acceptance by adult figures (Card 8), and conceptions of punishment (Card 10). A duplicate set of pictures corresponding to the CAT in style and composition was drawn by Heidi Schwabacher, graduate art student at the University of Minnesota. The duplicate pictures (Series A) contained human figures in place of animal figures; in all other respects the two sets of pictures were equivalent. Both picture series were ambiguous with respect to sex.

In order to counteract the facilitating effects of the initial test on the final test with respect to increased productivity, two groups of 30 children were chosen from the original sample by means of random numbers. Each group consisted of an equal number of boys and girls from each grade, thereby equalizing the factor of sex differences within the group. The two sets of pictures were administered in random order, using a test-retest method: Group A was given Series A, followed by Series B; Group B received Series B, followed by Series A. Retests were administered following a time interval of two weeks.

---

[1] The writer wishes to express her appreciation to Dr. Dale B. Harris for his helpful suggestions in the preparation of this manuscript.

[2] Now with State Department of Education, St. Paul.

Instructions to the subjects were held constant for both sets of pictures. The experimenter presented each picture to the child with the request, "Tell me a story about this picture." To clarify a point in the child's story the experimenter said, "Tell me more," "What do you mean?" or "Explain it to me." At appropriate places in the story the experimenter asked, "What happened before?" and "What will happen later?" The experimenter was limited to three questions in the administration of a single picture.

Measurement of the subject's reaction time was held constant for each picture. The child's response was timed with a stop watch from the presentation of the picture to the completion of the story. Total reaction time was recorded in seconds for each protocol.

Analysis of the fantasy data according to length of the response was done using the scoring system developed by McCarthy [2] in an earlier language study. Contractions of the subject and predicate counted as two words. Hyphenated words and compound nouns referring to single objects scored as one word. Contractions of the negative and the verb counted as a single word. Each part of the verb form was scored as one word.

Indices of the nouns, verbs, and ego-related words were computed on the basis of their frequency in the individual protocol. A word counted as a noun or verb depending on its function in the sentence. Compound nouns and verb combinations were considered single words. The phrases, "it looks like," "it seems as though," "it seems like," and the personal pronouns, I, my, me, and mine were scored in the Ego-Related Word Index. The response "I don't know" was not classified as an ego-related word.

Analysis of the stories in terms of nondescriptive verbalization was done using the transcendence categories outlined by Weisskopf [3]. The Transcendence Index of a picture was obtained by counting the number of comments about the picture which exceeded pure description. The mean number of nondescriptive remarks per subject constituted the index of the picture.

Two additional categories were used in computing a Transcendence Index, a sex-reference category and a naming category. The category of sex reference refers to the characterization of a figure in relation to its sex by the use of a singular or plural word. The naming category is the characterization of a figure by the use of proper nouns. The terms "baby" and "child" were not scored in any category.

The total number of seconds required by the subject to project himself into a particular picture was considered an index of reaction time.

Results

Means and standard deviations were obtained for length scores by sex and grade (Table 1). With grade progression, children tend to tell longer stories. Girls give slightly but consistently longer protocols than boys. There is no consistent tendency for animal pictures to evoke longer protocols than human pictures.

An analysis of nouns and verbs used, by the procedure employed in Table 1, showed slight but expected grade and sex differences, but no

Table 1

Length of Response Means and Standard Deviations by Grade and Sex for Human and Animal Pictures

Grade	Picture	Boys M	Boys *SD*	Girls M	Girls *SD*	Total M	Total *SD*
First	Human	61.48	29.25	64.54	36.45	63.01	33.08
	Animal	75.22	36.45	56.54	36.49	65.88	37.65
Second	Human	66.64	33.13	68.66	52.13	67.65	43.69
	Animal	64.38	33.02	70.80	35.77	67.59	34.57
Third	Human	62.30	44.33	86.02	46.09	74.16	46.75
	Animal	65.12	38.58	88.28	58.17	76.70	50.70

picture differences; the data are omitted in the interest of brevity.

Table 2 presents the means and standard deviations for ego-related word scores by grade and sex. A developmental trend may be disclosed in the grade means. The use of first person pronouns shows a slight but consistent decline with grade progression. Again, there are no differences attributable to picture source.

Table 2

Ego-Related Word Means and Standard Deviations by Grade and Sex for Human and Animal Pictures

Grade	Picture	Boys M	Boys *SD*	Girls M	Girls *SD*	Total M	Total *SD*
First	Human	1.92	1.41	2.64	1.73	2.28	1.73
	Animal	2.86	2.00	2.04	1.41	2.45	1.73
Second	Human	1.24	1.00	1.86	1.41	1.55	1.41
	Animal	1.64	1.73	1.62	1.73	1.63	1.73
Third	Human	1.62	1.41	1.46	1.41	1.54	1.41
	Animal	1.50	1.73	1.54	2.23	1.52	2.00

Mean transcendence scores and standard deviations by sex and grade are found in Table 3. The means for the three grades suggest a difference between the two series of pictures. Children give a greater number of nondescriptive comments in their responses to human figures than to animal figures. Girls tend to make more statements which are subjective and personalized than boys.

Table 3

Transcendence Means and Standard Deviations by Grade and Sex for Human and Animal Pictures

Grade	Picture	Boys M	Boys *SD*	Girls M	Girls *SD*	Total M	Total *SD*
First	Human	7.98	3.74	9.00	5.91	8.49	4.89
	Animal	7.42	3.74	6.86	4.79	7.14	4.24
Second	Human	9.30	3.74	9.50	6.16	9.40	5.09
	Animal	6.86	2.64	7.58	4.35	7.22	3.60
Third	Human	8.32	3.87	10.20	5.09	9.26	4.58
	Animal	7.12	4.24	9.50	6.24	8.31	5.47

Means and standard deviations were computed for reaction time scores by grade and sex. The data, omitted in the interest of brevity, show that girls have a consistently longer reaction time than boys in all grades.

Analysis of variance procedures were applied to the various scores reported in Tables 1 to 3 and in the text above. Table 4 summarizes the results of these various analyses. The sex differences conform to the well-known superiority of girls in various language measures. The grade differences, too, are not surprising. The results with respect to human and animal pictures are significant. Mean lengths of protocols and mean reaction times for the two sets do not differ. Nor do the sets differ in the number of naming responses, action words, or ego-related words. There is a significant difference between the animal and human figures with respect to Weisskopf's transcendence scores. Contrary to the assumption of the Children's Apperception Test, the difference favors the human pictures, which evoke significantly more subjective, personalized, interpretive responses, other than pure description. It is to be recalled that these conclusions are based on normal elementary school children, somewhat superior in intelligence. On every measure the variance among the means of children was significantly greater than the average variance within a child's responses to the several pictures.

Table 4

Results of Analyses of Variance Applied to Different Scores Derived from Thematic Picture Protocols

[Entries indicate the level of significance satisfied by the *F* ratio of the between-means variance tested against the remainder variance (the within-children variance)]

Comparison	Between means of sexes	Between means of grades	Between means of pictures	Between means of children
Length of protocol	.05	.01		.01
Number of nouns	.01	.01		.01
Number of verbs	.05			.01
Number of ego words		.01		.01
Transcendence scores	.01		.01	.01
Reaction time	.05			.01

Discussion

The original hypothesis that human pictures will evoke no more thematic material than animal pictures is refuted. The hypothesis that human pictures will evoke more nouns, verbs, and ego-related words fails to be refuted, as does the hypothesis that human pictures will induce more "blocking" as measured by reaction time to the picture.

Summary

Two duplicate series of pictures, one using animal and the other using human figures, were given to 60 children according to the following plan. Two groups of 30 children were chosen from the original sample by means of random numbers. Each group consisted of an equal number of boys and girls. Group A was tested with Picture Series A, followed by Picture Series B; Group B received Picture Series B, followed by Picture Series A. The picture series were separated by a time interval of two weeks. The arrangement of the pictures in a series for each subject was randomized by recourse to a table of random numbers.

1. With grade progression, children produce longer protocols. As children advance in grade level, they employ more nouns and fewer first person pronouns in their protocols.

2. Girls use a greater number of nouns and verbs in their protocols than boys. Girls employ more words and more nondescriptive comments in their protocols than boys. Girls have a slightly longer reaction time than boys.

3. Human figures evoke more verbalization which goes beyond pure description than do animal figures.

4. All scores used in this study are subject to greater variation among children than average variation "within children" (i.e., between all pictures for one child).

Received May 25, 1953.

References

1. Bellak, L. *Children's Apperception Test.* New York (P.O. Box 42, Gracie Station): C.P.S. Co., 1949.
2. McCarthy, Dorothea A. *The language development of the preschool child.* Minneapolis: Univer. of Minnesota Press, 1930.
3. Weisskopf, Edith A. A transcendence index as a proposed measure in the TAT. *J. Psychol.*, 1950, 29, 379-390.

Journal of Consulting Psychology
Vol. 18, No. 1, 1954

Delinquent vs. Nondelinquent Performance on the Porteus Qualitative Maze Test[1]

Richard F. Docter and C. L. Winder
Stanford University

Porteus [4] has described a "qualitative" method for scoring the Maze Test which appears to reflect differences in test behavior of delinquents and nondelinquents. This scoring scheme is based on an evaluation of the subject's regard for instructions, carefulness, impulsivity, etc. Essentially, it is a measure of expressive movements, and differs from the scoring of the mazes for intelligence ratings.

Studies by Porteus [2, 3], Wright [5], and Grajales [1] have revealed very significant differences between the qualitative performances of delinquents and nondelinquents. However, only Porteus has tested nondelinquents, the other authors having accepted his nondelinquent norms. (See Table 1.)

This study was conducted for the following two reasons: (*a*) to provide additional nondelinquent normative data, and (*b*) to provide an independent study of "matched" groups of delinquents and nondelinquents.

Procedure

Groups of 60 delinquent and 60 nondelinquent boys, matched for age, mental ability, race, and socioeconomic level, were given the Maze Test. The delinquents comprising this group were chosen for their similarity to those studied in previous investigations. All subjects were Caucasians, and were between the ages of 13 years, 10 months and 16 years, 10 months. Wechsler-Bellevue intelligence ratings for the delinquents ranged from 72 to 116. All delinquents were inmates in a state school for delinquent boys.

Nondelinquents were selected from the ninth grade of a junior high school and the tenth grade of a high school in San Jose, California. Their intelligence as measured by the California Test of Mental Maturity ranged from 71 to 128. In order to exclude noninstitutionalized delinquents from the nondelinquent group, the names of all subjects were reviewed by the school counselor and by the boys' physical education teachers. Only those who survived this process of elimination were tested. While the delinquent and nondelinquent groups are by no means perfectly matched, it appears certain that the most striking difference between groups is that one is characterized by "good citizenship," while the other is distinguished by its members' persistent disregard for the rules of society.

Results

The mean *Q* score (total weighted qualitative score) for delinquents was 47, which does not differ significantly from 49, the central tendency of previously reported means for delinquents (Table 1).

Table 1

Q-Score Means and Sigmas for Groups of Delinquents and Nondelinquents

Group		Delinquents			Nondelinquents		
		N	Mean	Sigma	*N*	Mean	Sigma
Porteus	[2]	100	49	25	100	22	13
Wright	[5]	54	49	28	—	—	—
Grajales	[1]	60	56	29	—	—	—
Porteus	[3]	50	48	22	179	19	13
Present study		60	47	26	60	25	20

The mean *Q* score for nondelinquents was 25, as compared to 22 which Porteus reported. This difference of three *Q*-score points is not significant. The major finding, then, is that the

---

[1] The writers express their gratitude to Robert Van Vorst and the staff of the Fred C. Nelles School for Boys at Whittier, California, and to personnel of the Public School System of San Jose, California.

qualitative performance of delinquents and nondelinquents corresponds almost exactly with the results previously reported for such groups, and that the difference between the delinquents' and nondelinquents' means yields a critical ratio in excess of five, which is significant at the .0001 level.

Table 2

Percentages of Delinquents and Nondelinquents Falling above Various *Q*-Score Critical Cutoffs

Group	Delinquents		Nondelinquents	
	Cutoff	Per cent above	Cutoff	Per cent above
Porteus [2]	29	80	29	21
Wright [5]	29	78	—	—
Grajales [1]	22	92	—	—
Present study	29	70	29	30

It is interesting to consider what predictive value this test might have. Porteus suggests a critical cutoff of 29 as the *Q* score which best discriminates delinquents from nondelinquents (Table 2). The writers found that of the delinquents, 70% fell above 29, while only 30% of the nondelinquents scored above this level. On the basis of these consistent differences between groups, this test shows promise of having predictive value.

Porteus probably constructed his weighting scheme after observing differences between the mean incidence of weighted errors for the two groups. Table 3 permits comparison of the means given by Porteus with the results of the present study, and it is noteworthy that the means reported by Grajales [1] and Wright [5] are very similar to these.

In order to assess the appropriateness of the Porteus weights, chi squares were computed based on differences in error incidence between the two groups. All the qualitative errors are discriminating in the expected direction (delinquents higher than nondelinquents), and on the basis of the magnitude of the respective chi squares, the Porteus weights appear to be reasonably assigned. Certain modifications could be justified on the basis of these data, but such changes would be premature.

That Porteus' weights are not essential is suggested by two findings. First, the critical ratio between the unweighted error means is 5.18, which is approximately the same as the critical ratio between the *Q*-score means. Second, using an unweighted error cutoff of 16, almost exactly the same percentages of each group fall above and below as when a *Q*-score cutoff of 29 is employed. It appears that the Porteus weights serve primarily to increase the magnitude of scores without altering the differences between groups. In the present study, the unweighted scores were: (*a*) delinquent mean $= 24.7$, $\sigma = 13.2$; (*b*) nondelinquent mean $= 13.7$, $\sigma = 9.9$.

Unfortunately, there are no data available on the reliability of this measure, but this study revealed a very high correlation between two scorers. Using the records of 60 nondelinquents, the product-moment correlation was .98; the rank-difference correlation, .96. Had the records of delinquents been used, these correlations might have dropped as the delinquents committed more errors which were difficult to score objectively; on the other hand, the increased range for the delinquents could compensate for th

Table 3

Mean Incidence of Qualitative Errors, the Porteus Weights, and Chi Squares for Differences between Delinquents' and Nondelinquents' Performance

Error	Mean incidence of qualitative errors				Porteus weights	Chi square	*p*
	Nondelinquents		Delinquents				
	Porteus [2]	Present study	Porteus [2]	Present study			
First third	1.01	0.60	1.58	1.63	2	4.76	.05
Last third	0.77	0.60	1.09	0.80	1	1.20	.30
Cut corner	3.01	3.15	4.40	4.61	1	7.00	.01
Cross line	3.18	3.50	5.93	7.68	2	13.04	.001
Lift pencil	1.63	2.80	5.77	3.81	3	2.40	.20
Wavy line	1.34	1.10	3.68	4.36	2	40.14	.001
Wrong direction	0.55	0.33	0.71	0.83	1	2.30	.20
Year V	0.27	0.28	1.09	0.46	1	2.84	.20
Year VI	0.33	0.21	0.99	0.50	1	2.48	.20

The extent to which *Q* score correlates with intelligence is a vital question. The results of the study corroborate previous reports which indicate the correlation is low. For delinquents who had taken the Wechsler-Bellevue, the *r* was —.27, significant at the .05 level. The nondelinquents' *Q* score correlated —.16 with the California Test of Mental Maturity, not significant. These should be interpreted as positive correlations, since a low *Q* score is "good."

In this research the entire series of mazes was given each subject, beginning with year V and continuing through the adult maze. The *Q* score was derived from this sequence, but the Maze Test intelligence rating was obtained only from that portion of this series which was applicable.

Since some subjects require two trials on the easier mazes and three or four trials to solve the more difficult ones, it might be asked what influence the number of mazes administered has upon the *Q* score. The range of mazes used was sufficient to justify the computation of a product-moment correlation between this variable and *Q* score. Combining both groups, the *r* was .29, which reaches the .01 level of significance. Therefore only about 9 per cent of the variance of *Q* score is attributable to variations in number of mazes used. It is recommended that the entire series, including the adult maze, be given.

Discussion

If one accepts the empirical demonstration that this test psychometrically distinguishes groups of adolescent boys whose social behavior is known to be different, there remains the question: What does the test measure? We confess that a justifiable answer for this question is not available at this time. But one guess which the writers suggest as reasonable is that this measurement of expressive movements taps one or more personality variables related to the construct of ego-control. Impulsivity may best describe the cluster of responses which contribute to the *Q* score. Should this be the case, and future research may clarify the matter, instruments of this kind would appear to be an important contribution toward the objective assessment of personality.

Summary

1. Groups of 60 delinquents and 60 nondelinquents were given the Porteus Qualitative Maze Test. The mean total weighted qualitative score (*Q* score) for delinquents was 47; for nondelinquents, 25. The difference was significant at the .0001 level.

2. Seventy per cent of the delinquents scored above 29; only 30 per cent of the nondelinquents exceeded this cutoff. The test shows promise of having predictive value.

3. Porteus' weighting scheme can be rationalized and used, but does not seem to add to the efficiency of the test.

4. A correlation of .98 showed reliability of scoring can be very high.

5. For delinquents, *Q* score correlated —.27 with intelligence, significant at the .05 level. For nondelinquents the correlation was —.16, not significant.

6. *Q* score correlated .29 with the number of mazes used, significant at the .01 level.

7. This test may measure personality variables, such as impulsivity, related to the construct of ego-control.

Received April 6, 1953.

References

1. Grajales, M. C. Porteus' Qualitative Maze Test as a measure of delinquency. Unpublished master's thesis, Fordham Univer., 1945.
2. Porteus, S. D. Qualitative performance in the maze test. Vineland, N. J.: Smith Printing House, 1942.
3. Porteus, S. D. Q-scores, temperament and delinquency. *J. abnorm. soc. Psychol.,* 1945, **21,** 81-103.
4. Porteus, S. D. *The Porteus Maze Test and intelligence.* Palo Alto, Calif.: Pacific Books, 1950.
5. Wright, Clare. The qualitative performance of delinquent boys on the Porteus Maze Test. *J. consult. Psychol.,* 1944, 8, 24-26.

NEW BOOKS AND TESTS

Books

Brown, Judson S., Harlow, Harry F., Postman, Leo J., Nowlis, Vincent, Newcomb, Theodore M., & Mowrer, O. Hobart. *Current theory and research in motivation: a symposium.* Lincoln: Univer. of Nebraska Press, 1953. Pp. v + 193 (paper). $2.00.

You will want to read this one. The six able contributors to the symposium are at their best and are, in the order named, as predictably Hullian, experimental, perceptual, developmental, social, and two-factored as you might expect. Would the subject matter were as predictable as the protagonists! But they shed more light on the complex problem of motivation than has been shed in many a moon. If you are interested in psychological theory, send to Nebraska for your copy — now. — *L.F.S.*

Daniel, Robert S., & Louttit, C. M. *Professional problems in psychology.* New York: Prentice-Hall, 1953. Pp. xi + 416. $5.50.

This book supplies a sound core around which to build a professional problems course for graduate students in the applied fields of psychology. It contains a good assortment of materials relevant to the skills which psychologists must have in addition to their subject matter — about jobs, training, history, bibliography, written communication, publication, organizations, ethics, legislation, and the like. The style is clear and graceful but mature; it does not sidestep complex and difficult issues by talking down to the reader. The graduate student, very soon to be a member of the profession, is recognized as interested in and capable of understanding the affairs of his field. The book integrates well with basic source materials such as the APA's *Publication Manual* and *Ethical Standards of Psychologists*. It is in agreement with these documents and supplements them without undesirable duplication. Now that this good resource is available there should be more and better orientation courses for our students. And many already-graduated psychologists may read it with enjoyment and profit, too. — *L.F.S.*

Doll, Edgar A. *The measurement of social competence: a manual for the Vineland Social Maturity Scale.* Philadelphia: Educational Test Bureau, 1953. Pp. xviii + 664. $7.75.

The Vineland Social Maturity Scale is an unusual psychological instrument, not a test but a method for quantifying observations of social maturation. It is one of the few devices that can be applied reliably and usefully in the absence of the examinee from the reports of persons who have observed his developmental achievements. The Scale, available since 1936, is now provided with a comprehensive manual which enhances its usefulness. The manual describes the conception and construction of the scale, its administration and scoring, and its standardization and validation. The author draws both from his own basic studies and from the published research of others. Most of the evidence supports the value of the Scale: its good repeat reliability, its effective scaling of items, and its applicability to many practical and research problems. It is something of a shock, however, to find that the present standardization is based on only 20 cases, 10 male and 10 female, at each age from 0-1 to 30. The paucity of cases is offset by the unusual care exercised in selecting the normative subjects and by the supporting data from consecutive ages. The manual, like the Scale, has many merits and a few clear shortcomings. The completeness of the data presented, and the cautiousness of the interpretations, may serve as models for other test authors. On the negative side, the manual, while attractive in appearance, is organized so as to make essential data hard to find; the index seems to provide frustration as often as help. The diligent searcher, however, is ultimately rewarded. The Social Maturity Scale has now come to its own maturity, and should become as essential as a clinical tool as the intelligence test. — *L.F.S.*

Dollard, John, Auld, Frank Jr., & White, Alice Marsden. *Steps in psychotherapy.* New York: Macmillan, 1953. Pp. ix + 222. $3.50.

This is an exciting little book. In simple language, the authors undertake to orient the reader to the Dollard-Miller theory of therapy, "a blend of reinforcement learning theory and psychoanalysis," and then proceed to demonstrate the implications of the theory for practice by presenting selected portions of the seventeen recorded therapy sessions constitut-

---

Note. — The reviews were prepared by the Editor and the Associate Editors, who may be identified by their initials.

ing brief psychotherapy for a 40-year-old married woman patient. Discussion of these materials is largely in terms of the recorded comments of the therapy supervisor (Dollard) to the young therapist (an anonymous psychiatrist) during the supervisory sessions. Part II includes the complete protocols of the battery of tests administered between the first and second therapy hours and two independent sets of test interpretations by clinical psychologists, one written with and the other without access to a summary of the first therapy session. The authors do not claim validity for either the therapeutic or diagnostic procedures but rather offer these raw materials and their ideas as promising approaches to the development of better knowledge. The book should be extremely useful as a basis for seminar discussions and of interest to all psychologists concerned with the training of psychotherapists. — *E.L.K.*

Eysenck, H. J. *The structure of human personality.* London: Methuen, and New York: Wiley, 1953. Pp. xix + 348. $5.75.

Eysenck's most recent contribution to the study of personality is a survey of research. But it is more than a survey; it is a critical and creative integration which brings order to a diversity of quantitative studies. Eysenck espouses factor analysis as the prime technique for revealing the organization of personality. In an introductory chapter on theory he identifies first-order factors with "traits" and second-order factors with "types." The research evidence is reviewed in a series of chapters arranged according to the technique of appraisal employed: ratings, questionnaires, objective tests, measurements of physique, physiological indices, attitude and interest measurements, and correlations between persons. Many studies are reinterpreted, sometimes with new analyses of their data, so as to bring a semblance of agreement to results previously considered contradictory. Eysenck does not indulge in the petty virtue of impartiality but the blows he gives his opponents, while often severe, are fair and well aimed. The book will appeal to psychologists and students who prefer theory buttressed by research to vague professions of faith. They may not agree with every word but they will learn and be stimulated. — *L.F.S.*

Festinger, Leon, & Katz, Daniel. (Eds.) *Research methods in the behavioral sciences.* New York: Dryden, 1953. Pp. ix +660. $5.90.

No cookbook of investigative techniques, this stimulating volume represents a serious and vigorous attempt to clarify the logic of a number of research problems in the sciences of individual and collective human behavior. Such methods as the sample survey, the field study, and the formal experiment under both field and laboratory conditions are analyzed in terms of their advantages and disadvantages with respect to various substantive issues in social and psychological theory. Two chapters are devoted to the technical problems of sample selection and distribution-free statistical procedures, both of which should prove especially meaningful for research workers in clinical psychology and in the psychology of personality. One section deals with techniques of observing group behavior, of using documents and census records, and of collecting usable data by interview. On the analytic side, there are contributions on the use of qualitative data and on the theory and methods of social measurement. While it is a little unfair in a book of such remarkably even and high quality to give special mention to particular sections, Helen Peak's chapter on the general problem of making objective observations could profitably be read by every graduate student in the country. Similarly, the concluding chapter by Likert and Lippitt on the utilization of social science is an insightful discussion, at once seminal and sound, of the neglected problem of how research results can be most fruitfully capitalized upon both by other researchers and by applied workers and citizen leaders. This book will richly repay serious and repeated consideration by psychologists of very different theoretical persuasions and specific investigative interests. — *E.J.S.*

Hathaway, Starke R., & Monachesi, Elio D. (Eds.) *Analyzing and predicting juvenile delinquency with the MMPI.* Minneapolis: Univer. of Minnesota Press, 1953. Pp. viii + 153. $3.50.

A research monograph containing six studies based on the complete testing of a ninth-grade group with the MMPI in 1947-48, and a subsequent comparison with police and court records. Moderate relationships between some MMPI indicators and delinquency were found which can be interpreted with the aid of expectancy tables. The most valid prediction was the nondelinquency of youngsters with clearly normal profiles. A valuable side product of the study is a useful amount of data about adolescents in general. — *L.F.S.*

Kinsey, Alfred C., Pomeroy, Wardell B., Martin, Clyde E., & Gebhard, Paul H. *Sexual behavior in the human female.* Philadelphia: Saunders, 1953. Pp. xxx + 842. $8.00.

There will probably be as many lines of type set about this book as there are in it, which should make a reviewer cautious lest he review its stimulus value rather than the volume per se. The book is similar in many respects to the 1948 report on the male. It differs in that it does not contain the large part on "factors affecting"; instead, a part is devoted to a comparison of the sexes. The Kinsey team admits it has not undertaken "U. S. corrections" of the raw data. However, there is a point-for-point comparison with male results, which raises questions about the significance of these generalizations. The authors frankly admit the inadequacy of their sample, especially for females who have not gone beyond grade school. Female sexual behavior is studied with orgasm and erotic arousal as the dependent variables, with age, education, marital status, occupation, religious and rural-urban backgrounds, age of onset of adolescence, and decade of birth as independent variables. In the male study mere sexual arousal was not included as a variable. Religious background and decade of birth were conspicuously important variables; age, a very important variable in male outlet, was related differently to female behavior. Males were found to be conditioned by sexual experience more frequently than were females, and they share vicariously in the sexual experiences of others to a greater extent. The data for youth show majority behavior following somewhat the course dictated by cultural pressures. Kinsey and his associates did more than report findings. They suggested social implications which go beyond the scope of their molecular study. Whereas the authors generally seemed to show scientific caution and scholarliness, some of the earlier criticism of the limitations of their method does not seem to have influenced the second study greatly. — *F. McK.*

Lehman, Harvey C. *Age and achievement.* Princeton, N. J.: Princeton Univer. Press, 1953. Pp. xiii + 359. $7.50.

Psychologists will be grateful for this volume which brings together for the first time Lehman's penetrating studies of age and achievement, previously published in journal articles over an interval of nearly twenty years. Like many good empiricists, Lehman raises more questions than he answers. Why is outstanding creativity achieved earlier in chemistry than in the social sciences? Why do our creative scientists seem to be getting younger, while our statesmen are getting older? These facts, and many others as interesting, are attested by Lehman's careful studies. The "whys" remain vital topics for additional investigation. — *L.F.S.*

Peck, Leigh. *Child psychology.* Boston: Heath, 1953. Pp. 536. $5.25.

Child Psychology is planned for college students on the undergraduate level. The author has constantly kept in mind the special needs of students preparing to work with children, and has given a comprehensive picture of psychological development. It is refreshing to find a book that makes its subject matter such delightful reading for the group for which it is planned. Pertinent reports of actual situations illustrate the discussions and help the reader clarify his own thinking. The arrangement, organization, references, and index are excellent and the book should prove helpful in an introductory course in child development. — *B.M.L.*

Piaget, Jean. *The child's conception of number.* New York: Humanities Press, 1952. Pp. ix + 248. $5.00.

This is a translation of a book previously published in French some twelve years ago. Piaget's naturalistic approach to the mind of the child is here applied to his development of the number function. As usual, Piaget has included material from the original protocols, and the book makes interesting reading for the psychologist. The practicing clinician will get more insight here into the functioning of the child's mind in a testing situation than he will from the abstractions of factor analysis. The translation is good and the format pleasing. — *W.A.H.*

Powdermaker, Florence B., & Frank, Jerome D. *Group psychotherapy.* Cambridge: Harvard Univer. Press, 1953. Pp. xv + 615. $6.50.

This is a detailed report of two experiments in group therapy. Both studies were conducted under the auspices of the Washington School of Psychiatry, one with 124 VA clinic patients, mostly neurotics, and the other with 88 VA hospital patients, mostly schizophrenics. Only for the psychotic group was a control group used. The magnitude of the experiment can be seen from the fact that 24 "medical doctors" were involved (10 of them psychiatrists), 5 clinical psychologists, and 2 psychiatric social workers. There were 12 groups in the clinic experiment and 15 in the hospital. In the neurotic group, 453 sessions were studied. The major method employed was the situational analysis of group meetings, developed from Harry Stack Sullivan's concept of situations as bases of human behavior. The analyses make excellent teaching (and learning) material. An equally valuable portion is an appendix with a running account of the first and the 65th meeting of one group. Situations are also presented to illustrate processes in the conduct of group therapy. One aspect of these studies will probably evoke some controversy. Fourteen of the 24 M.D.'s apparently had had no previous clinical training, but they were permitted to do "psychotherapy" with their groups. In fact, according to the authors, their lack of experience turned out to be an advantage. On the other hand, the clinical psychologists and psychiatric social workers were not employed in the group therapy. One is tempted to ask some questions about the reasoning involved in these decisions. However, the importance of this volume lies in the discussions of research methods in group psychotherapy, rather than in the results obtained by the therapy. — *M.K.*

Sargent, Helen D. *The Insight Test.* Menninger Clin. Monogr. Ser. No. 10. New York: Grune & Stratton, 1953. Pp. xii + 276. $6.75.

Sargent has supplied a more complete manual for her ingenious pencil-and-paper projective method which was first described in 1944. In the test, subjects' written responses to described conflict situations are scored for expressed affect (*A*), defense (*D*), and malignancy (*M*) and for several subscores under each of these main categories. Much use is made of the ratio *A/D*. The manual gives the theory underlying the test, its administration, scoring, and interpretation, and copious illustrative protocols drawn from seven clinical groups. When judged by the standards of a quantitative test, the method receives meager support. Although the scoring reliability of trained scorers is good, the retest reliability of subjects is reported only for inadequate groups and even then ranges from poor to mediocre. The "normal" group, on which standards are based, consists of 20 screened university students. Obviously, evidence that patients differ from such "normals" has no value. The author gives evidence of being a competent statistician, and discusses the needs for controls of age, sex, education, and socioeconomic status but does nothing about them. A fair inference is that she does not really believe in experimental evaluation. The test, of course, can be judged by other criteria. The clinical protocols reveal it as a rich source of fertile hunches about psychopathology. Until better research evidence is produced, however, the *Insight Test* must remain a better device for making hypotheses than for testing them. — *L.F.S.*

Viteles, Morris S. *Motivation and morale in industry.* New York: Norton, 1953. Pp. xvi + 510. $9.50.

This volume, written primarily for management, "is devoted to a comprehensive description and critical evaluation of American and British experimental studies and attitude surveys bearing upon the sources of motivation and the determinants of morale in industry." It is indeed comprehensive and the bibliographies at the end of each chapter are extensive. Some background in motivational theories is given. The problem of mobilizing the will-to-work is discussed with particular reference to wage-incentive systems. Studies on the effect on productivity of such variables as job preference, size of work group, knowledge of results, supervision, and employee participation are covered effectively. There is also a section on employee attitude surveys and a final one on guideposts for management, with cautions on application of such surveys and the other techniques discussed. — *A.R.*

Tests

Finch, F. H., Kuhlmann, F., & Betts, G. L. *Kuhlmann-Finch Intelligence Tests.* Kgn.-gr. 12, 9 levels. 1 form, each level. 25 (30) min. Test booklet ($1.95 per 25) with instructions, keys; IBM answer sheet (95¢ per 25); manual, pp. v + 57 ($1.00); specimen set ($1.50). Philadelphia: Educational Test Bureau, 1953.

The new series of tests bears only a superficial resemblance to its predecessor, the Kuhlmann-Anderson. The tests consist of entirely new items planned by Kuhlmann and developed after his death by Finch. Each booklet has five subtests with lenient five-minute timing. Nonverbal tests are used through grade 3, and two verbal subtests, two nonverbal, and one numerical at the higher levels. MA and IQ scores are obtained from charts in the booklets, with little clerical effort. The manual gives only 10 pages to the construction and standardization of the test, and there are no references to research articles. While not extensive, the technical data give a decidedly favorable impression. Items were selected not only to yield progressive age increments, but to be as free as possible of cultural and sex biases. Reliabilities are good (.86 — .92); norms are based on large and widespread samples. Data on validities, except against the general developmental criterion, are still lacking. — *L.F.S.*

King, Joseph E., & Wingert, Judith W. *IPI Merit Rating Series.* Industrial employees. 5 scales, clerical, mechanical, sales, supervisor, technical. Rating form ($4.00 per 20); manuals, pp. 20, 8; specimen set ($3.00). Tucson, Ariz. (Box 6056): Industrial Psychology, Inc., 1953.

The *Merit Rating Series* applies a psychological test concept to the rating of employee performance. In place of the usual rating scale, 60 statements are provided in the form for each job family. The rater marks each statement *yes* or *no*, and weighted scoring yields quantitative scores for several performance traits and for over-all competence. The manuals discuss the methods of rating, training of raters, and application of results. At least minimally acceptable data are given on item difficulties and consistencies, the derivation of weights, and norms. In addition to supervisory uses, the method seems to be of potential value as a criterion for studies of personnel aptitude and selection. — *L.F.S.*

Books Received

Abrahams, Joseph, & Varon, Edith. *Maternal dependency and schizophrenia.* New York: International Universities Press, 1953. Pp. 240. $4.00.

Bonaparte, Marie. *Female sexuality.* New York: International Universities Press, 1953. Pp. x + 225. $4.50.

Cottle, Wm. C. *The MMPI: a review.* Lawrence, Kan.: Univer. of Kansas Press, 1953. Pp. vi + 82.

Donahue, Wilma, Rae, James Jr., & Berry, Roger B. (Eds.) *Rehabilitation of the older worker.* Ann Arbor: Univer. of Michigan Press, 1953. Pp. v + 200. $3.25.

Ginzberg, Eli, Herma, John L., & Ginsburg, Sol W. *Psychiatry and military manpower policy.* New York: King's Crown Press, 1953. Pp. xi + 66. $2.00.

Greenacre, Phyllis. (Ed.) *Affective disorders.* New York: International Universities Press, 1953. Pp. 212. $3.00.

Heston, Joseph C. *How to take a test.* Chicago: Science Research Associates, 1953, Pp. 47. 40¢.

Ingram, Christine P. *Education of the slow-learning child.* New York: Ronald, 1953. Pp. ix + 359. $5.00.

Kahn, Alfred. *A court for children.* New York: Columbia Univer. Press, 1953. Pp. xxii + 359. $4.50.

Leitner, Konradi. *Hypnotism for professionals.* New York: Stravon, 1953, Pp. 127. $4.00.

Little, Wilson, & Chapman, A. L. *Developmental guidance in secondary school.* New York: McGraw-Hill, 1953. Pp. ix + 324. $4.50.

Nichtenhauser, Adolf, Coleman, Marie L., & Ruhe, David S. *Films in psychiatry, psychology and mental health.* New York: Health Education Council, 1953. Pp. 269. $6.00.

Prosser, Charles A., & Sifferd, Calvin, S. *Selecting an occupation.* Bloomington, Ill.: McKnight & McKnight, 1953. Pp. x + 246. $2.50.

Ralli, Elaine P. (Ed.) *Adrenal cortex* (Transactions of the Fourth Conference, 1952). New York: Josiah Macy, Jr. Foundation, 1953. Pp. 165.

Reik, Theodor. *The haunting melody.* New York: Farrar, Straus, & Young, 1953. Pp. viii + 376. $4.00.

Sartre, Jean-Paul. *Existential psychoanalysis.* New York: Philosophical Library, 1953. Pp. viii + 275. $4.75.

Serog, Max. *New light on dreams.* Boston: House of Edinboro, 1953. Pp. 159. $3.00.

Tyler, Leona E. *The work of the counselor.* New York: Appleton-Century-Crofts, 1953. Pp. x + 323. $3.00.

Vernon, Philip E. *Personality tests and assessments.* London: Methuen, 1953. Pp. ix + 220. $4.00.

Whitaker, Carl A., & Malone, Thomas P. *The roots of psychotherapy.* New York: Blakiston, 1953. Pp. xvii + 236. $4.50.

Wortis, Joseph. (Ed.) *Basic problems in psychiatry.* New York: Grune & Stratton, 1953. Pp. vi + 186. $4.50.

Yost, Orin Ross. *What you should know about mental illness.* New York: Exposition Press, 195 Pp. 165. $3.50.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

Prognostic Factors in Schizophrenia[1]

William Schofield, Starke R. Hathaway, Donald W. Hastings,
and Dorothy M. Bell
University of Minnesota

Vol. 18 No. 2 Meeting

In view of the magnitude of the problem presented by the psychiatric enigma known as schizophrenia, it is startling to find how relatively little careful attention has been paid to the question of outcomes in this illness and how few comprehensive investigations have been made into prognostic factors. A reasonably exhaustive review of the research literature on outcome and prognosis in schizophrenia reveals only 31 published studies in this country between 1909 and 1951 — less than one report per year. These 31 studies cover a total of only 6,802 patients (exclusive of Pollock's [7] statistical review of 19, 927 patients), the different samples ranging in size from 20 to 1,082 patients. The length of the follow-up interval has ranged from zero to fifteen years, with an average of approximately four and one-half years in those studies which clearly specify the length of follow-up. Of the 31 studies reported, fewer than one-third have used direct face-to-face contacts with the patients as a basis for the follow-up evaluations, the remainder relying on questionnaire responses and letters. In view of the extent of the variability in the factors of sample size, length of follow-up period, and basis of follow-up evaluation, it is not surprising that the figures reported for "recovery" rate range from 0 to 35%, and that those for "improvement" rate extend from 5 to 74%!

A survey of the literature reveals that true investigations into prognosis, i.e., attempts to ascertain factors differentiating patients with good outcomes from those with poor, constitute a very small proportion of the follow-up studies, the majority of which are simple reports of the status of a given sample of patients at some period following hospital admission or discharge. The worker wishing to review specifically prognostic studies labors under some difficulty because of the fairly widespread tendency of investigators to use the term "prognosis" in studies which are purely statistical reports on "outcomes," with no effort to analyze factors related to outcome.

No attempt will be made in this paper to present a detailed review of the literature summarized above. A comprehensive review of the published studies on outcome and prognosis in schizophrenia, with a detailed bibliography, has been prepared for separate publication; the European researches are included in this review. However, it is appropriate here to call attention to a few papers which have particular import. In 1938, Romano and Ebaugh [8] published one of the earliest of the few studies in which both questionnaire responses and the results of direct re-examination were available for the same sample of patients as bases on which to appraise their postdischarge status. Comparison of the results of direct examination with questionnaire data from the families showed significant error in over 50% of the cases for which both types of data were available.

The paper by Malamud and Render [6] is important on a number of counts. It is one of the few studies that recognizes the marked discrepancies among the "recovery and improvement" statistics previously reported, and it reviews the possible sources of these discrepancies. It is concerned with an actual attempt to distinguish between schizophrenics with poor and good outcomes, respectively. It is based on a sample of over 300 patients and evaluates their status after an average postadmission interval of six years. Most important of all, the follow-up evaluation was based on direct re-examination; and all patients had been originally examined by one of the investigators.

---

[1] This study was made possible by the Frederick B. Wells, Jr. Trust Fund. A condensation of this paper was read at the meeting of the American Psychological Association, Washington, September, 1952.

In 1941, Chase and Silverman [3] reported a "critical survey of the literature" on prognostic criteria in schizophrenia. This was by no means a comprehensive review, since the reference list of 49 items failed to include 15 of the 30 follow-up studies published in the United States through 1940. Thus, the degree of agreement or disagreement as to the prognostic value of particular items might actually prove quite different from that indicated by Chase and Silverman in their limited survey. Discussing the findings of various studies under 15 different factors, they include that definite prognostic value attaches to the factors of duration of illness, acuteness of onset, presence of obvious exogenic precipitants, confusion and atypical symptoms, "process" symptoms, and subtypes. The survey suggested questionable or no prognostic validity for the factors of age, sex, education, psychosexual history, previous episodes, body type, extroversion-introversion, and heredity. While this review is not exhaustive, it does help to indicate the very limited number of factors on which the research studies are uniform in their findings.

One of the most thorough reviews of the literature on prognosis was published by Blair [1] in 1940. This is particularly good with respect to the European studies, although it is less complete in its coverage of American research. It is primarily concerned with "spontaneous remission" rates as a base line for evaluating efficacy of cardiozol and insulin therapies.

Purpose and Subjects

The purpose of the present investigation was to add to the very limited research data on prognostic factors in schizophrenia. It was a further object of the study to avoid some of the errors and sources of inadequacy of previous studies by using a reasonably large sample of patients, by evaluating their *social adjustment* through direct contact after an interval of some years following initial hospitalization, and by attempting a reasonably exhaustive study of case histories and clinical data in a search for prognostic indicators.

The criterion data on which this study is based were collected as part of an over-all follow-up study of psychiatric patients [5]. In 1949, a graduate social worker trained in research traveled throughout Minnesota to interview patients who had been hospitalized on the psychiatric service of the University of Minnesota Hospitals between 1938-1944. This interval was chosen because it covered the period from the inception of the psychiatric service to the start of extensive utilization of electric and convulsive therapies. Thus, data on the subsequent adjustments of patients evaluated in this study permit appraisal of the effects of psychiatric hospitalization exclusive of the shock therapies which have assumed a major role in recent years.

The social worker succeeded in establishing contact with the patient directly, or with close relatives, friends, and neighbors, or with well-informed agencies in the home community, in 77% (1,261) of the total of 1,638 cases treated during the 1938-1944 interval. In terms of the frequency of the various diagnostic categories, the 77% sample was found to be essentially identical with the total original group of hospitalized patients. Schizophrenics constituted 14.1% of the original group and 16.7% of the group with which follow-up contact was achieved. Of the 16 diagnostic groups represented in the total follow-up sample, the schizophrenics ranked second for percentage of follow-up contact; 87.5% of the 240 patients hospitalized in the interval under study were traced by the social worker. General descriptive data for these patients are reported in Table 2.

A word should be said concerning the nature of the treatment received by the schizophrenic patients. They were hospitalized for an average of 40 days ($M = 39.9$; $\sigma = 30.1$) on a small locked ward (35 beds). The psychiatric service was primarily a diagnostic and teaching unit with a small, eclectically oriented staff. In the assignment of formal diagnoses the official nosological system of the American Psychiatric Association was applied. The usual medical facilities of a psychiatric hospital were available and were used as indicated (e.g., sedatives, hydrotherapy, nursing management of regimen, etc.). Occupational therapy facilities were available and were used extensively in the treatment program; general recreational facilities were extremely limited. Because of the small staff and the generally diagnostic orientation of the unit, psychotherapy was of a most limited nature. No use was made during the interval under study of convulsive or "shock" therapies. It might reasonably be held that recovery and improvement figures for this sample would reflect relatively "spontaneous" remission, and that factors found to differentiate between those with good and those with poor posthospital courses might be used to equate later patient groups for study of the efficacy of specific therapies such as insulin shock.

Procedures

On the basis of interviews and hospital records for those cases in which rehospitalization had occurred, the social worker assigned a rating to each patient to indicate the degree of difficulty he had experienced in his over-all adjustment subsequent to his hospitalization at the university. A five-step rating scale was used, with each step defined as follows:

A. Patient had no further trouble of the type for which he was hospitalized originally.

B. Patients were largely making a good social adjustment. To a small extent they had been bothered by the complaints they presented at the time of hospitalization, but these occurred very infrequently, bothered the patient or his family to a minimal degree, or had gradually diminished and disappeared within the year following hospitalization. These patients may have spent a short time in a state hospital; usually it was immediately following initial hospitalization at the university, and the patient was soon paroled and never returned.

C. Patients were in and out of state hospitals, or worked and made a good adjustment for a time but had relapses of the illness for which they were hospitalized. Although these patients had been well for relatively long periods of time, they usually had two or three episodes of illness.

D. Patients had spent more than half of the time between university hospitalization and follow-up in an institution, incapacitated because of the illness, or functioning on the inadequate level which characterized them at the time of university hospitalization. For the most part, the patient or the patient's family considered him to have a large amount of trouble.

E. Patients had continuous trouble of the type for which they were hospitalized. Many of them were sent from the university hospitals directly to a state institution, and they remained there. Others were cared for at home by their families, but none of them recovered even for a short while.

Table 1 presents the distributions of the follow-up ratings for the total sample of 210 schizophrenics contacted. In comparing the figures of this table with those in the existing literature, it should be noted that the criterion of evaluation applied in this study was not one of degree of psychiatric symptomatology, but rather one of *social adjustment,* and that the interval of posthospital adjustment ranged from five to eleven years.

Schizophrenic patients who fell into the four extreme categories of rated posthospital adjustment (A and B, D and E) constituted the criterion groups in a search for prognostic factors. Adequate case history data and clinical notes were available for all of the A and B cases and for all but seven of the D and E cases. The criterion groups were thus composed of 58 patients who had no further difficulty of the type for which they were originally hospitalized or who were making a good social adjustment (Good group), and 119 schizophrenics who had spent over 50% of the time since their university hospitalization in a mental institution or who had experienced continuous difficulty necessitating their being cared for by their families (Poor group).

The hospital records and case histories of the 178 schizophrenics were next analyzed, with reference to a very detailed schedule for notation of presence or absence, or a rating of the degree, of some 250 items covering personal history, medical history, and development and features of the psychosis.[2] A random selection of records was independently rated on the case history variables by two raters. Complete agreement as to rating was found to obtain on between 80% and 95% of the items in indi-

---

[2] Copies of the 12-page rating schedule for case history analysis and of the 4-page record form used with it have been deposited with the American Documentation Institute. Order Document No. 4209 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. S., remitting in advance $1.25 for microfilm or $1.25 for 6 x 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

This schedule was originally developed by W. Grant Dahlstrom in connection with an unpublished doctor's dissertation [4].

Table 1

Rated Posthospital Adjustment of Schizophrenics Who Received No Specific Form of Therapy

Adjustment rating	N	% of group
A (best)	14	7
B	44	21
C	26	12
D	29	14
E (worst)	97	46
Total	210	100

vidual records. Free use was made of an "unknown" category in order not to introduce invalidity by forcing ratings on variables for which data were frequently inadequate. Ratings of the case histories were made without knowledge of which records belonged to which outcome group.

The case histories provided adequate data for a sufficient number of patients to permit statistical analyses of the differential appearance of 200 items in the two criterion groups. Basic data derived from the four parts of the case history analyses are reported in Tables 2-5. While it is clear from inspection that much of the data reported in these tables has no prognostic significance, it is presented in detail as normative statistics which describe the manner in which the clinical material of schizophrenia presents itself. It is hoped that these data may serve as an addition to the textbook descriptions of this psychosis, which generally are unsupported by such surveys.

Of the 200 items statistically tested, only 17 met the 5% level of reliability. While this number of positive findings, and their content, is such as to argue against mere chance [2], the fact that only 17 distinctions should be yielded by 200 items attests to the subtlety of the social and personal factors determining postpsychotic adjustment.

Results

Table 2 presents general descriptive data for the groups with good and with poor posthospital adjustment, and for the two groups combined. The age variable is age at the time of hospitalization. The Good group has both a lower mean age at hospitalization and a reliably more restricted age range than the Poor group ($F = 1.553$; $p < .01$). As is to be expected, these differences are paralleled by the data on "age at onset" (Table 3). The Good group is younger at age of onset and has a reliably smaller variance on this variable than the Poor group ($F = 2.293$; $p < .01$).

As further shown in Table 2, the two outcome groups are *not* differentiated by factors of sex, urban versus rural home, education, religion, marital status, number of children, and number of siblings. While the Good group is characterized by nearly twice the proportion of catatonics as the Poor group, the distribution of the various subtypes in the two outcome groups is not reliably different ($\chi^2 = 9.12$; $p < .10$).

Of the thirty items of personal history recorded in Table 3, only four discriminate between the two groups. The Poor group is characterized by a reliably higher frequency of *excellent* deportment in school ($\chi^2 = 7.82$; $p < .02$). The Good group reveals a reliably higher frequency of church attendance ($\chi^2 = 9.67$; $p < .01$). The marital adjustment of the Good group is clearly superior to that of the Poor group, showing no incidence of "extreme frustration" or "continual conflict," and nearly three times as many "pleasant" marital relationships are recorded for the Good group as for the Poor. Affection toward the mate was manifested by nearly all of the married schizophrenics with good outcome (94.1%), and by only 50% of those in the Poor group ($t = 4.17$; $p < .01$).

In view of some of the prominent thinking concerning psychogenic etiology, it is of considerable interest to note from Table 3 that the two groups are not different with respect to interparental relationships, relationship with either parent, relationship to sibs, or sibling rank (Items 1-5). It is also interesting to note that they are not differentiated by the extent of social participation (Item 13), by their heterosexual adjustment (Item 19), or by their initiative, social intelligence, and poise (Items 22, e, f, and g).

Table 4 reviews the data descriptive of the presenting illnesses of the schizophrenic patients. As noted above, the patients with good outcome were characterized by a reliably smaller range of age at onset (Item 7). The dura-

Table 2

Gross Descriptive Data for Schizophrenics with Good and Poor Posthospital Courses *

Datum	Good group $N = 58$	Poor group $N = 119$	Total group $N = 177$
1. Age	$M = 25.2$ $SD = 7.0$	$M = 30.5$ $SD = 10.0$	$M = 28.7$ $SD = 9.6$
2. Sex			
Male	47.4%	42.8%	44.4%
Female	52.6	57.2	55.6
3. Home			
Urban	41.3%	55.4%	50.8%
Rural	58.7	44.6	49.2
4. Education	($N = 57$)	($N = 110$)	($N = 167$)
Less than 8 gr.	12.3%	10.0%	10.8%
8 gr. & more	87.7	90.0	89.2
5. Religion			
Lutheran	39.1%	33.7%	35.7%
Other Prot.	23.9	34.9	31.0
Catholic	34.8	22.8	27.1
Jewish	2.2	8.6	6.2
6. Number of Sibs	$M = 4.4$ $SD = 3.2$	$M = 4.1$ $SD = 3.4$	$M = 4.2$ $SD = 3.4$
7. Marital status			
Married	27.1%	26.8%	26.9%
Single	71.2	68.0	69.1
Divorced	0.0	2.5	1.7
Widowed	1.7	1.6	1.7
Separated	0.0	1.1	0.6
8. Number of children	($N = 17$) $M = 2.2$ $SD = 2.2$	($N = 39$) $M = 3.1$ $SD = 2.4$	($N = 56$) $M = 2.8$ $SD = 2.4$
9. Type of schizophrenia			
Catatonic	33.8%	18.4%	23.6%
Simple	18.6	21.8	20.8
Paranoid	16.9	35.2	29.2
Hebephrenic	10.1	9.2	9.6
Unclassified or mixed	20.6	15.4	16.8

* The N's in this and subsequent tables vary with different item comparisons in terms of the number of patients for whom there was adequate information. Where N's are not indicated for the item, data are for the total groups, i.e., $N = 58$, $N = 119$, respectively.

tion of the illness prior to hospitalization was considerably shorter in the Good group than in the Poor group (Item 8, $\chi^2 = 22.2$; $p < .01$). The rate of a history of previous episodes was twice as great in the Poor group as in the Good group (Item 9, $t = 2.83$; $p < .01$). The distribution of the "poor" and "fair" categories of clinical prognosis at time of discharge yields a small but reliable separation of the two outcome groups (Item 11; $\chi^2 = 4.36$; $p < .05$). The Good group had a reliably smaller variation in length of hospitalization than the Poor group (Item 12 *a*, $F = 1.53$; $p < .05$). Finally, the adjustment to hospitalization of the Good group was inferior to that of the Poor group (Item 12 *b*, $\chi^2 = 9.58$; $p < .01$). It is of particular interest to note from Item 12 *c* of Table 4 that the long-term adjustment of the schizophrenics was not reliably reflected in their condition at time of discharge; nearly two-thirds of the patients who later made a good adjustment were clinically judged as unimproved when discharged, while 20% of the patients with subsequently poor social adjust-

Table 3

Personal History Data for Schizophrenics with Good and Poor Posthospital Courses

Item	Good group N = 58	Poor group N = 119	Total group N = 177
1. Interparental relat.	(N = 35)	(N = 66)	(N = 101)
Affection	82.8%	72.7%	76.2%
Ambivalence	5.7	6.0	5.9
Indifference	0.0	3.0	1.9
Hostility	11.4	18.2	15.8
2. Paternal relat.	(N = 41)	(N = 86)	(N = 127)
Affection	70.7%	62.8%	65.3%
Ambivalence	9.8	3.4	5.5
Indifference	0.0	5.8	3.1
Rejection	9.8	12.8	11.8
Overprotection	0.0	2.3	1.5
Domination	9.8	12.8	11.8
Jealousy	0.0	0.0	0.0
Dependency	0.0	0.0	0.0
Neglect	0.0	0.0	0.0
3. Maternal relat.	(N = 47)	(N = 81)	(N = 128)
Affection	61.7%	66.6%	64.8%
Ambivalence	2.1	2.4	2.3
Indifference	4.2	0.0	1.5
Rejection	2.1	8.6	6.2
Overprotection	12.8	13.6	13.2
Domination	14.8	8.6	10.9
Jealousy	0.0	0.0	0.0
Dependency	0.0	0.0	0.0
Neglect	2.1	0.0	0.7
4. Sibling relat.	(N = 43)	(N = 83)	(N = 126)
None	4.6%	3.6%	3.9%
Affection	76.7	75.9	76.1
Indifference	4.6	4.8	4.7
Rivalry	11.6	14.4	13.4
Domination	0.0	1.2	0.7
Submission	2.3	0.0	0.7
5. Sibling rank	(N = 34)	(N = 63)	(N = 97)
0	5.8%	4.8%	5.1%
1	26.4	28.6	27.8
2	20.6	19.0	19.5
3	17.6	14.2	15.4
4	14.7	17.4	16.5
5	5.8	6.3	6.1
6	5.8	3.2	4.1
7	0.0	1.6	1.0
8	2.9	0.0	1.0
9	0.0	0.0	0.0
10	0.0	3.2	2.1
11	0.0	1.6	1.0
6. Home			
Poverty	5.0%	10.9%	9.0%
Alcoholism	6.8	6.7	6.8
Invalid	1.6	0.0	0.6
Divorce	1.6	1.6	1.6
Separation	3.4	0.0	1.1
Death of parent	10.2	10.9	10.7
7. School acceptance	(N = 53)	(N = 106)	(N = 159)
Marked hostility	5.8%	2.9%	3.8%
Mild dislike	9.4	12.2	11.3
Indifference	24.5	26.4	25.8
Agreeable	52.8	55.6	54.7
Keen enjoyment	7.5	2.9	4.4
8. School achievement	(N = 56)	(N = 107)	(N = 163)
Repeated failure	7.3%	3.7%	4.9%
Work difficult	14.2	22.4	19.6
Av. performance	58.9	45.8	50.2
Easily earned grades	10.7	24.4	19.6
Accelerated	8.9	3.7	4.4
9. School deportment	(N = 37)	(N = 77)	(N = 114)
Poor record	13.6%	5.2%	7.9%
Usual no. of escapades	21.6	5.2	10.6
Excellent record	64.8	89.6	81.6
10. Occupation			
None	22.0%	17.6%	21.1%
11. Success	(N = 40)	(N = 87)	(N = 127)
Poor	7.5%	18.4%	14.9%
Average	30.0	32.2	31.5
Good	62.5	49.4	53.5
12. Satisfaction	(N = 40)	(N = 86)	(N = 126)
Dislike	12.5%	8.1%	9.5%
Indifference	32.5	48.8	43.6
Enjoyment	55.0	43.0	46.8
13. Social	(N = 31)	(N = 54)	(N = 85)
Withdrawal	58.1%	63.0%	61.2%
Ambivalence	12.9	14.8	14.1
Membership	29.0	22.2	24.7
14. Dating	(N = 45)	(N = 100)	(N = 145)
None or little	60.0%	56.0%	57.2%
Average	40.0	44.0	42.8
Very popular	0.0	0.0	0.0
Steady dating	35.6	36.1	35.9
15. Recreation	(N = 55)	(N = 108)	(N = 163)
Predominantly solitary	43.6%	38.8%	40.5%
Mixed	47.2	54.6	52.2
Social	9.2	6.4	7.3
16. Church attendance	(N = 49)	(N = 98)	(N = 147)
Very infrequent	16.3%	41.8%	38.3%
Occasionally	42.8	26.5	31.2
Steady	40.8	31.7	35.3
17. Delinquency	3.4%	2.5%	2.8%
18. Criminal record	5.0%	5.0%	5.0%
19. Heterosexual adjust.	(N = 58)	(N = 113)	(N = 171)
Poor	22.4%	22.1%	22.2%
Fair	31.0	34.5	33.3
Good	46.6	43.4	44.4
20. Adequacy of outlet	(N = 58)	(N = 112)	(N = 170)
Poor	17.2%	28.6%	24.7%
Fair	32.8	29.4	30.5
Good	50.0	42.0	44.7
21. Marital			
None	71.2%	71.4%	71.3%
a. Adjustment	(N = 16)	(N = 34)	(N = 50)
Extreme frustration	0.0%	23.5%	16.0%
Continual conflict	0.0	20.6	14.0
Compatibility	50.0	38.2	42.0
Pleasant	43.8	14.7	24.0
Chief pleasure	6.2	2.9	4.0
b. Affection to mate	(N = 17) 94.1%	(N = 34) 50.0%	(N = 51) 64.7%
22. Personality			
a. Breadth of interest	(N = 58)	(N = 114)	(N = 172)
Narrow	27.1%	39.4%	35.2%
Some interest outside self	64.4	57.8	60.2
Broad interests	8.5	2.8	4.6
b. Level of aspiration	(N = 58)	(N = 112)	(N = 170)
Limited goals	34.4%	34.8%	34.7%
Interested in improving	56.8	54.4	55.3
High ambition	8.8	10.8	10.0
c. Life plan	(N = 54)	(N = 110)	(N = 164)
Vague	44.4%	47.2%	46.4%
Confused and unrealistic	30.0	27.2	28.0
Fairly clear	25.6	25.6	25.6
d. Stability	(N = 58)	(N = 114)	(N = 172)
Constant fluct.	16.9%	12.2%	13.9%
Moderate variability	42.3	42.9	42.7
Stolid, even tempered	40.6	44.7	43.4
e. Initiative	(N = 58)	(N = 111)	(N = 169)
Apathetic	13.8%	16.9%	16.0%
Appropriate	62.6	48.6	53.2
Energetic	23.6	39.5	30.8
f. Social intelligence	(N = 58)	(N = 109)	(N = 167)
Inept and clumsy	3.4%	11.9%	8.9%
Moderate skill	93.2	86.2	88.8
Adept,	3.4	1.9	2.3
g. Poise	(N = 58)	(N = 112)	(N = 170)
Retiring and sensitive	37.2%	38.4%	37.4%
Fairly articulate	50.8	48.2	49.7
Confident in most situations	11.8	13.4	12.8
h. Religiousness	(N = 43)	(N = 78)	(N = 121)
Largely intellectualized and ritualistic	0.0%	6.4%	4.1%
Occasional solace	58.2	53.8	55.4
Dominant source of balance and peace of mind	41.8	39.7	40.5

Table 4

Elements in the Presenting Illness of Schizophrenics with Good and Poor Posthospital Courses

Item	Good group $N=58$		Poor group $N=119$		Total group $N=177$	
1. Areas of manifestation of illness	% Patients	% Problems	% Patients	% Problems	% Patients	% Problems
Home	35.2	18.2	45.6	19.6	37.1	19.2
Social	20.4	10.6	26.2	11.2	21.4	11.1
Occupational	25.9	13.4	35.9	15.4	28.7	14.8
Religious	3.7	1.9	6.8	2.9	5.1	2.6
Sexual	16.6	8.6	31.0	13.3	23.0	11.9
Personality	90.6	47.1	87.3	37.4	87.1	40.4
2. General health	($N=54$)		($N=108$)		($N=162$)	
Invalid	0.0%		1.9%		1.2%	
Many minor illnesses	13.0		13.0		13.0	
Usual diseases	18.5		21.2		20.4	
Occasional ailment	40.7		28.7		32.7	
Very good health	27.8		35.2		32.7	
3. Handicaps	6.7%		2.5%		3.9%	
4. Deformities	1.6%		1.6%		1.6%	
5. Weight	($N=45$)		($N=79$)		($N=124$)	
Underweight	2.2%		5.1%		4.1%	
Thin	11.1		19.0		16.1	
Well-proportioned	82.2		67.1		72.5	
Buxom	0.0		6.3		4.1	
Overweight	4.4		2.5		3.2	
6. System involvement						
Cardiorespiratory	3.3%		0.8%		1.7%	
Gastrointestinal	1.6		2.5		2.2	
Genitourinary	1.6		5.8		4.5	
7. Age at onset	($N=58$) $M=24.6$ $SD=7.0$		($N=114$) $M=26.4$ $SD=10.6$		($N=172$) $M=25.9$ $SD=8.7$	
8. Development of symptoms	($N=58$)		($N=116$)		($N=174$)	
Over a year	25.8%		62.0%		50.0%	
A month to a year	55.2		30.2		38.5	
Less than a month	19.0		7.8		11.4	
9. Previous episodes	18.6%		37.8%		31.4%	
10. Etiology	($N=55$)		($N=116$)		($N=171$)	
Personality	98.1%		99.1%		98.8%	
Family	9.0		11.2		10.5	
Social	10.9		5.2		7.0	
Occupational	10.9		2.6		5.2	
Capacity	3.6		4.3		4.0	
Health	5.4		6.8		6.4	
11. Prognosis	($N=51$)		($N=110$)		($N=161$)	
Poor	43.1%		60.9%		55.2%	
Fair	52.9		36.3		41.6	
Good	3.9		2.7		3.2	
12. Course in hospital						
a. Length	($N=57$) $M=38.6$ days $SD=27.6$		($N=119$) $M=40.4$ days $SD=34.4$		($N=176$) $M=39.9$ days $SD=30.1$	
b. Adjustment	($N=53$)		($N=108$)		($N=161$)	
Poor	28.2%		13.0%		18.1%	
Fair	49.0		42.6		44.6	
Good	22.8		44.4		37.3	
c. Progress	($N=56$)		($N=118$)		($N=174$)	
No change	62.5%		66.9%		65.5%	
Small improv.	10.7		16.9		14.9	
Decided improv.	26.8		20.3		22.4	

Table 5

Mental Status of Schizophrenics with Good and Poor Posthospital Courses

Item	Good group N = 58	Poor group N = 119	Total group N = 177
1. Delusion	(N = 22)	(N = 44)	(N = 66)
Mild transitory	36.4%	31.8%	33.3%
More extensive	50.0	45.4	47.0
Complex, well-established	13.6	22.7	19.7
Grandeur	3.4%	0.8%	1.7%
Influence	8.4	9.2	9.0
Nihilism	5.0	1.6	2.8
Persecution	35.6	36.1	35.9
Reference	15.2	27.7	23.6
Religion	13.6	14.2	14.0
Self-change	1.6	3.4	2.8
Sex	13.6	5.8	8.5
Somatic	32.2	26.8	28.6
Unreality	5.0	1.6	2.8
2. Hallucination			
Auditory	35.6%	31.0%	32.6%
Visual	32.2	22.6	25.8
Olfactory	1.7	2.6	2.2
Haptic	0.0	0.8	0.6
Gustatory	0.0	0.0	0.0
3. Complaints—hypochondriacal	3.4%	7.6%	6.2%
4. Feelings			
Apathy	15.2%	17.6%	16.9%
Depersonalization	3.4	6.7	5.6
Doubt	0.0	0.8	0.6
Guilt	18.6	16.0	16.9
Inferiority	1.6	4.2	3.4
Isolation	1.6	0.0	0.6
Perplexity	10.2	3.4	5.6
Unreality	6.8	7.6	7.3
5. Orientation			
a. Time	(N = 51)	(N = 110)	(N = 161)
Gross disorientation	19.6%	6.5%	10.6%
Minor errors	9.8	7.2	8.1
Answers well	70.6	86.3	81.3
b. Place	(N = 51)	(N = 111)	(N = 162)
Gross disorientation	9.8%	3.6%	5.6%
Minor errors	9.8	2.8	4.9
Answers well	80.4	93.6	89.5
c. Person	(N = 50)	(N = 111)	(N = 161)
Gross disorientation	8.0%	4.5%	5.6%
Minor errors	4.0	3.6	3.7
Answers well	88.0	91.9	90.7
6. Memory			
a. Past	(N = 46)	(N = 107)	(N = 153)
Vague	6.5%	4.7%	5.2%
Minor errors	6.5	9.3	8.5
Accurate	87.0	86.0	86.3
b. Recent events	(N = 46)	(N = 101)	(N = 147)
Vague	15.3%	9.0%	10.9%
Minor errors	21.7	12.8	15.6
Accurate	63.0	78.2	72.5
c. Recent verbal	(N = 27)	(N = 44)	(N = 71)
Vague	25.9%	22.7%	24.0%
Minor errors	18.5	9.1	12.5
Accurate	55.6	68.2	63.5
d. Recent numerical	(N = 36)	(N = 69)	(N = 105)
Vague	16.6%	11.7%	13.3%
Minor errors	27.8	27.5	27.6
Accurate	55.6	60.8	59.0
7. Retrospective falsification	(N = 9) 44.0%	(N = 21) 28.5%	(N = 30) 33.3%
8. Thinking capacity	(N = 42)	(N = 98)	(N = 140)
Low	9.5%	9.2%	9.3%
Below average	35.7	36.7	36.4
Average	47.6	41.8	43.5
Above average	7.2	9.2	8.6
High	0.0	3.1	2.1
9. General information	(N = 41)	(N = 93)	(N = 134)
Low	14.6%	5.4%	8.2%
Below average	14.6	34.4	28.3
Average	51.2	43.0	45.6
Above average	14.6	14.0	14.2
High	5.0	3.2	3.7

Table 5 (Continued)

10. Emotional tone	(N = 57)	(N = 114)	(N = 171)
a. Stability			
wide lability	14.0%	8.8%	10.5%
change	14.0	24.6	21.0
stable	72.0	66.6	68.4
b. Depression-Elation	(N = 54)	(N = 112)	(N = 166)
Depressed	55.5%	28.5%	37.3%
Composed	31.5	60.6	51.2
Happy	13.0	10.9	11.9
c. Anxiety	23.7	16.0	18.5
d. Moodiness	3.4	3.4	3.4
e. Irritability	10.2	5.0	6.8
11. Insight	(N = 57)	(N = 117)	(N = 174)
Poor	57.8%	69.2%	65.5%
Fair	40.4	29.9	33.3
Good	1.8	0.9	1.1
12. Stream of thought			
a. Refusal-Flight	(N = 58)	(N = 115)	(N = 173)
Refusal to speak	12.0%	6.0%	8.1%
Retardation	24.1	21.7	22.5
Quiescent	24.1	20.8	22.0
Normal	12.0	27.0	22.0
Verbose	7.0	5.3	5.8
Voluble	7.0	9.6	8.6
Flight of ideas	13.8	9.6	11.0
b. Consist.-Distract.	(N = 51)	(N = 108)	(N = 159)
Consistent	66.6%	61.1%	62.9%
Circumstantial	9.9	10.2	10.0
Rambling	23.5	28.7	27.1
c. Coher.-Incoher.	(N = 53)	(N = 107)	(N = 160)
Coherent statements	90.6%	89.7%	90.0%
d. Impoverishment	6.8	4.2	5.1
e. Echolalia	1.6	0.0	0.6
13. Attitude, manner, behavior			
a. Cooperativeness	(N = 57)	(N = 116)	(N = 173)
Poor	26.3%	14.6%	18.5%
Fair	22.8	32.8	29.5
Good	50.8	52.6	52.0
b. Rapport	(N = 54)	(N = 116)	(N = 170)
Superficial	42.6%	33.6%	36.5%
Friendly	42.6	40.5	41.2
Close	15.8	25.9	22.3
c. Psychomotor			
Stupor	6.8%	2.5%	4.9%
Apathy	50.8	50.4	50.5
Suggestibility	1.6	0.8	1.1
Mimicry	0.0	0.8	0.6
Facetiousness	3.4	0.8	1.7
Querulousness	0.6	2.5	1.7
Hostility	16.9	5.8	9.6
Assaultiveness	11.8	8.4	9.6
d. Neatness	(N = 49)	(N = 103)	(N = 152)
Dishevelled	30.6%	26.2%	27.6%
Neglected	24.6	29.1	27.6
Well-groomed	44.8	44.6	44.7
e. Maturity	(N = 49)	(N = 45)	(N = 94)
Immature	6.9%	0.0%	2.7%
Minor anachronism	0.0	2.2	1.3
Behavior mature	93.1	97.8	96.0
f. Miscellaneous			
Panic	5.0%	2.5%	3.4%
Tension	22.0	8.4	12.9
Tremor	5.0	3.4	3.9
Blushing	1.6	0.0	0.6
Tics	1.6	1.6	1.6
Mannerisms	11.8	18.4	16.3
Hyperkinesis	13.6	11.8	12.4
Catalepsy	10.2	7.6	8.5
14. Symptoms			
a. Affect. cont. Disharmony	(N = 42)	(N = 93)	(N = 135)
Appropriate display of emot.	50.0%	43.0%	45.1%
Mild dislocation	40.4	40.8	40.7
Gross splitting	9.6	16.2	14.0
b. Phobias	1.6	4.2	3.3
c. Insomnia	8.4	7.6	7.9
d. Hypersensitive	1.6	1.6	1.6
e. Rigidity	6.8	2.5	3.9
f. Echopraxia	3.4	0.0	1.1

Table 5 (Continued)

g. Stereotypy	15.2	0.8	5.6
h. Litigious	3.4	2.5	2.8
i. Jealousy	1.6	0.8	1.1
j. Fatigue	1.6	2.5	2.2
k. Dreamy states	5.1	1.6	2.8
l. Night terrors	5.1	0.0	1.9
m. Compulsions	3.4	0.8	1.9
n. Egocentric	3.4	5.0	4.5
o. Temper tantrum	3.4	4.2	3.9
p. Amorality	0.0	0.8	0.6
q. Regression	5.1	5.0	5.1
r. Exhibition	3.4	2.5	2.8
s. Indecision	3.4	1.6	2.2
t. Autism	0.0	0.0	0.0
15. Other behavior			
Inappropriate	18.6%	16.0%	16.9%
Silly	11.9	7.6	9.0
Suicidal	5.0	4.2	4.5
Stubborn	23.7	25.2	24.7
Illogical	3.4	2.5	2.8
Suspicious	16.9	23.5	21.3
Impulsive	3.4	5.8	5.0
Childish	5.0	5.0	5.0
Affected	0.0	0.8	0.6
Remorseful	1.6	1.6	1.6
Self-accusatory	5.0	3.4	3.9
Tearful	20.3	6.7	11.3
Nail biting	3.4	1.6	2.2
Insecure	5.0	1.6	2.8
Self-deprec.	3.4	6.7	5.6

ment were considered to have made marked psychiatric improvement.

Over 90 different items pertaining to the mental status and symptomatology of the schizophrenic patients are reported in Table 5. Only six of these items show reliably different frequencies in the two outcome groups. Ideas of reference (Item 6) occurred in a reliably larger proportion of the Poor group ($t = 2.00$; $p < .05$). The Good group showed nearly three times the incidence of gross time disorientation (Item 15*a*) was was found in the Poor group ($\chi^2 = 7.68$; $p < .05$). With respect to emotional tone (Item 20*b*), the Good group was much more frequently described as dull or depressed and much less frequently described as composed when compared with the Poor group ($\chi^2 = 13.27$; $p < .01$). Tension (Item 23*f*) was reliably more frequent in the Good group ($t = 2.30$; $p < .05$.). Stereotypy (Item 24*g*) was practically absent in the Poor group, but noted in 15% of the patients with good post-hospital adjustments ($t = 3.054$; $p < .01$). Finally, tearfulness (Item 25*b*) was over three times as frequent among those schizophrenics with good adjustments as it was in those with poor ($t = 2.40$; $p < .05$).

Some attention should be paid to the failure of all but this very limited number of mental status items to distinguish the two groups. In particular, note is made of the fact that the degree of delusional involvement was not differentially distributed in the two groups, that feelings of apathy had essentially equal incidence, and that appropriate display of affect characterized nearly half of each subgroup. It is equally notable that persecutory, religious, and somatic delusions; degree of insight; and inappropriate, stubborn, and suspicious behavior do not distinguish the two groups from one another. Of much interest, also, is the very low incidence in both groups of symptoms which are commonly considered to be classically descriptive of schizophrenia. For example, in the total sample of 177 schizophrenics, there was a less than 5% occurrence of nihilistic delusions and delusions of self-change and unreality; olfactory, haptic, and gustatory hallucinations; echolalia; stupor, suggestibility, and mimicry; and temper tantrums, exhibitionism, and autism.

A prognostic scale was devised consisting of eight items selected as the most discriminating of the seventeen items reviewed above. The eight items, each of which was given an unweighted unit score in a direction predictive of poor prognosis, were as follows:

1. School deportment: "excellent record; avoided all trouble."
2. Marital adjustment: "source of extreme frustration and emotional turmoil," or "continued mild bickering and conflict."
3. Time taken for symptoms to develop: over a year or more.
4. Presence of previous episodes.
5. Adjustment to hospital: "good; cooperated, showed initiative, carried out assignments well."
6. Presence of ideas of reference.
7. Orientation for time: "answers well within the limits of approximation afforded by sources of information."
8. Emotional tone: "complacent or composed."

Application of this scale to the 177 patients included in the study, using a cutting score of three or more points to identify the cases with poor adjustment, resulted in identification of 59% of the cases in the Good group and 82% of the cases in the Poor group. These figures should be evaluated against the actual 29% occurrence of good outcomes and the 58% incidence of poor outcomes. The degree of improvement over a chance prediction indicated by these figures does not constitute adequate

Table 6

Means and Standard Deviations for the 8-Point Prognostic Scale Applied to Three Different Outcome Groups

Posthospital Adjustment	*N*	*M*	*σ*
Groups A and B	58	2.11	1.24
Group C	26	2.68	1.05
Groups D and E	119	3.68	1.58
Total	203	3.12	1.60

validation of the prognostic scale, however, since the figures represent application of the scale to the same group of subjects on which it was developed. Table 6 gives data on the distribution of the prognosis scale scores in the three outcome groups for subjects whose histories permitted scoring of the eight factors. The statistics of the intermediate group (Group C), which was not used in development of the prognosis scale, satisfy the logical requirements which pertain if there is a reliable, positive relationship between the prognosis scale scores and the criterion of actual posthospital adjustment; i.e., the mean score is intermediate to that of the two extreme groups, and the standard deviation is less than that of either of the two other groups (A and B, D and E) which constitute larger segments of the criterion continuum.

Attention is called to Item 11 of Table 4, which reports the accuracy of clinical prediction of posthospital adjustment. The psychiatric staff appears to have been more pessimistic about this illness in the sample of patients involved than the follow-up data would justify. While nearly one-fourth of the total group had good posthospital courses, the staff predicted only 4% of such outcomes. Furthermore, they predicted poor subsequent histories for over 40% of the patients who were found actually to have achieved good social adjustment. It must be considered that the prognostications of the psychiatric staff may have been more heavily weighted with considerations of degree of continuing psychiatric involvement (symptomatology) rather than with the less technical, but socially crucial, question of adjustment to the community. Nevertheless, the need for factual information as to prognostic factors is clearly illustrated.

Discussion

The findings of this study may be compared with the conclusions reached by Chase and Silverman [3] as a result of a *partial* survey of the literature on prognosis in schizophrenia through 1940. They found short duration of illness, acuteness of onset, presence of confusion in the early symptomatology, and manic-depressive symptoms to be indicative of good prognosis. The findings of the present research are in agreement on these points. Duration of illness in our data is reflected in the relative discrepancies between the ages at onset and ages at hospitalization in the groups with good and with poor outcomes respectively. The onset-hospitalization interval averages .6 years in the Good group and 4.1 years in the Poor group. Acuteness of onset is reflected in the "rate of development" variable (Table 4) and shows slow development of symptomatology to characterize the Poor group. Confusion is indicated in the disorientation for time, which was more common in the Good group, and manic-depressive symptomatology is suggested by the greater occurrence of depression and tearfulness in the Good group.

Chase and Silverman indicate as also prognostically significant, but less differentiating than the above factors, the fact of previous episodes, adequate prepsychotic life adjustment, and progressive hospital course. The findings of this study are directly *opposite* on these points. Thus, occurrence of previous episodes appears as more characteristic of the Poor group; most of the variables pertaining to previous adjustment (*vide* Table 3, Items 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19) do not discriminate the Good from the Poor group, and course in the hospital (Item 12*c*) showed no reliable differences between the two outcome groups.

With respect to failure of sex, intelligence, education, psychosexual history, and age at onset per se to show prognostic significance, this study agrees with the conclusions of the Chase and Silverman [3] survey.

The present research has considered a number of factors not generally covered in previous investigations. There is a set of factors relatively social in nature, or at least reflective of interpersonal relationships, which differentiates the two outcome groups. In brief, the Good group was characterized by poor school deportment, steady church attendance, good marital adjustment, and affection toward mates. General intelligence as evaluated during the mental status examination revealed no marked difference between the two groups, and hence the difference in school deportment appears unrelated to a difference in ability.

What seems at first glance to be a strange inconsistency in the adjustment of the Good group to the two highly organized institutions responsible for the development of social conformity, namely, the school and the church, is possibly accounted for by the relative demand for active participation by the individual as an individual. The seclusiveness and shyness which would handicap the schizophrenic in the schoolroom would be no detriment to his largely spectator role in church. This explanation, of course, presumes typical prepsychotic personality features and is open to argument on that basis. It is further weakened by the fact that in the present study "social withdrawal," as rated on the basis of case histories, revealed no differential frequency in the two groups! The fact of positive association between good marital adjustment with affection toward the mate, and good posthospital course, seems most reasonable.

It may be pointed out that the poorer hospital adjustment of the members of the group with good outcome is probably related to their tendency toward briefer hospitalization, both of these factors suggesting more acute disturbance in the Good group; this would be in line with usual prognostic thinking and with the greater frequency of the catatonic type in the Good group. To some extent it may be assumed that the greater incidence of stereotypy, tension, and time disorientation in the group with good posthospital adjustment is associated with the greater incidence of the catatonic subtype in this group. Similarly, the occurrence of previous episodes and ideas of reference as more characteristic of the Poor group may reflect the slight predominance of the paranoid type in this group. However, it should be recalled that the over-all distribution of the subtypes was not different for the two outcome groups, and the prognostic factors that have been reviewed in detail may not be considered to simply stem from subtype variation.

The validity of the findings of this study rests heavily on the thoroughness and accuracy of the case histories, which are functions of the skill and care with which they were compiled by the medical, psychological, and social work personnel. It is perhaps unfortunate that these were routine and not "research" histories and that they were not recorded with the guidance of the schedule of items subsequently applied to their analysis. However, the essential casualness of the history taking, in addition to the fact that items were rated only when the data were considered adequate, may introduce a type of clinical validity and permit a generalization of our prognostic factors that might not be the case if the original history data were more "scientific." Quite positively, it is believed that the nature and extent of the follow-up study have provided more accurate and socially pertinent criteria of posthospital adjustment than have previously been applied in studies of prognosis.

The paucity of careful follow-up studies involving reasonably large samples, extended postdischarge intervals, and direct evaluation of subsequent adjustments of patients may account for the generally pessimistic orientation which is held with respect to prognosis in schizophrenia. This pessimism is reflected in the clinical prognoses which were assigned to the patients in this study and appears to have been unjustified in view of their subsequent social adjustments. It is hoped that data of the type presented in this paper may contribute to greater exactness in prognosis, not only by indicating the actual frequency of favorable and unfavorable long-term outcomes, but also by suggesting historical and clinical factors which carry prognostic import. Certainly, good information as to basic rates of remission in the absence of special therapeutic efforts is essential to any reasonable test of the efficacy of particular therapies, such as insulin coma. In establishing control and experimental groups for testing therapeutic efficacy, attention must be given to adequate matching on variables, affecting basic prognosis. As an example of this, discrepancies in the remission rates reported for samples of schizophrenics similarly treated might be explained in some instances by marked dissimilarity of the samples with respect to diagnostic subtype composition, if these data were reported by the investigators.

Summary and Conclusions

In 1949, a graduate social worker with special research training and experience attempted to establish direct contact with all patients hospitalized on the psychiatric service of the University of Minnesota Hospitals between 1938-1944, an interval when no use was made of insulin or electroshock therapy. Of the original sample of 240 schizophrenics, follow-up data were obtained on 210 patients. These were rated on a five-step scale for their over-all posthospital social adjustment. Twenty-eight per cent of this sample were found to have had no further difficulty or were found to be making a good adjustment (Good group), and 60% were found to have spent over half of the post-University interval in a mental institution or had experienced difficulty which necessitated continuous care by their families (Poor group).

The Good and the Poor outcome groups were rated and contrasted with respect to 200 items pertaining to their general history, previous personality, and presenting illness. Only 17 of the 200 items differentiated the two groups at the 5% level of probability. This number of positive findings from such a large series of tests is sufficient to reject the null hypothesis with respect to the frequency of the items in the populations represented by the two samples. However, the smallness of the number of differentiating items is much more impressive than the fact of statistical reliability. To the degree that the posthospital adjustments of schizophrenics are determined by factors of a historical, personal, and social nature, these factors would appear to be of an extremely subtle nature. It would appear that discovery of those factors in the structure of family relationships — the physical aspects of the home environment, and the neighborhood and community attitudes toward the returned patient — that may be important in determining the quality of his adjustment are highly elusive, and their detection would require a much more intensive clinical evaluation of the posthospital course than has as yet been achieved. The other possibility suggested by the findings of this study would be that subsequent social adjustment of the schizophrenic is determined by factors of a largely endogenous nature that are inherent in the disease process.

With respect to prognostic indicators, the patients with good posthospital adjustments were differentiated from those with poor adjustments by the factors of: shorter duration of illness prior to hospitalization; more rapid development of symptomatology; absence of previous episodes; poor school deportment; steady church attendance; good marital adjustment and affection toward the mate; stereotypy; tension; depression; tearfulness; disorientation for time; and absence of ideas of reference.

Received October 2, 1953.

References

1. Blair, D. Prognosis in schizophrenia. *J. ment. Sci.*, 1940, **86**, 378-477.
2. Brozek, J., & Tiede, K. Reliable and questionable significance in a series of statistical tests. *Psychol. Bull.*, 1952, **49**, 339-341.
3. Chase, L. S., & Silverman, S. Prognostic criteria in schizophrenia: a critical survey of the literature. *Amer. J. Psychiat.*, 1941, **98**, 360-368.
4. Dahlstrom, W. G. An exploration of mental status syndromes by factor analytic techniques. Unpublished doctor's dissertation, Univer. of Minnesota, 1949.
5. Hastings, D. W., Hathaway, S. R., & Bell, Dorothy. Follow-up study of psychiatric patients. *Univer. Minn. Hosp. Bull.*, 1951, **23**, 149-159.
6. Malamud, W., & Render, N. Course and prognosis in schizophrenia. *Amer. J. Psychiat.*, 1939, **95**, 1039-1057.
7. Pollock, H. M. Outcome of mental diseases in the United States. *Ment. Hyg., N. Y.*, 1925, **9**, 783-804.
8. Romano, J., & Ebaugh, F. G. Prognosis in schizophrenia; a preliminary report. *Amer. J. Psychiat.*, 1938, **95**, 583-596.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

Direction of Aggression as a Prognostic Variable in Mental Illness

Dorothy A. Feldman
University of Pittsburgh

G. R. Pascal, and C. H. Swensen
University of Tennessee

The differential prognosis of paranoid and depressive disorders has a long history in psychiatric literature [2]. However, predictions based on these diagnostic categories have not been very reliable. In some instances, paranoids do improve and depressed patients do not [5, 8]. That the current classificatory system is unreliable is generally conceded [2, pp. 86-87], and the need for variables which are reliable and which have prognostic value for the outcome of treatment is clear. Taking its cue from the classical differentiation between paranoid and depressive disorders [3] and from studies by Albee [1] and Smith [9], the present study considers direction of aggression as a parameter of mental illness and its relation to prognosis. As part of a larger research project [6], this study also attempts to avoid one of the major weaknesses of earlier prognostic studies in that, as far as possible, other variables of known prognostic significance are held constant while direction of aggression is studied.

Method and Procedure

A previous paper described the selection of a sample of 486 patients from the population of hospitalized mental patients at the Western Psychiatric Institute and Clinic [6]. There were 229 males and 257 females in the sample. Diagnostic categories were represented as follows: 264 with schizophrenia, 58 with affective psychoses, 17 with "other psychoses," 95 with psychoneuroses, and 52 with behavior disorders. Status was categorized as improved or unimproved, according to follow-up information one year after discharge. There were 234 improved and 252 unimproved individuals. Every practical effort was made to insure the reliability of the "improved-unimproved" dichotomy, and evidence was presented to indicate the validity of the judgments [6]. The sample contained no subjects younger than 15 or older than 55 years and no known cases of cortical damage or mental deficiency. No patients had a serious somatic illness, and none had a history of recovery from previous alienations.

Guided by the literature and the feasibility of obtaining reasonably reliable measurements, we selected 21 variables and related them, by appropriate statistical techniques, to status one year after discharge from the hospital. Of these 21 variables, the following 9 were found to be significantly related to the defined outcome of treatment: affective expression, orientation, diagnosis, onset, duration of illness, precipitating stress, type of treatment, marital status, and direction of aggression. The intercorrelations of these variables were also investigated. Then an attempt was made to hold eight of the relevant variables constant while investigating the relationship between status and the ninth variable, direction of aggression. In other words, the group of improved and unimproved patients were of the same size, contained the same number of diagnostic categories, and had no statistically significant differences between their means on length of illness, type of onset, stress, and so on. Such matching, which was deemed best for greater precision, reduced the original sample of 486 to a controlled sample of 100 improved and 100 unimproved subjects.

Direction of aggression was measured by a nine-point rating scale, details of which are described elsewhere [4]. It was found to have a reliability coefficient of .83 when two clinical psychologists independently rated 40 case histories selected at random from the hospital files. The nine points along the continuum from extremely extrapunitive to extremely intropunitive behavior are presented below.

Point one: homicidal attack. The use or threatened use of lethal weapons. Attempts to strangle or maim. Blame projected onto others in conjunction with this behavior.

Point two: combative behavior. Assaultive, extremely destructive, while projecting blame. This point may include a previous suicidal attempt which was obviously aggressive in nature, followed by assaultive behavior.

Point three: verbal abusiveness, or systematized paranoid delusions expressed with or without verbal abuse.

Point four: blame projected onto physical complaints, as long as the hypochondriasis is not bizarre or depersonalized. Vague, unsystematized projection. Ambivalent behavior, with occasional aggressive activity toward others.

Point five: ambivalent, impunitive. This point includes obsessive-compulsive behavior and incoherent states.

Point six: mild depression. Ambivalent behavior with some evidence of depression, as in apathy with occasional tears. Bizarre hypochondriacal complaints.

Point seven: depressed. Feelings of guilt, self-blame expressed, with no projection of blame.

Point eight: deeply depressed. Continuous verbal abuse of the self, expressing sin, guilt, unworthiness. This point may include expression of suicidal wishes and suicidal attempts of a milder type than in point nine.

Point nine: serious suicidal attempts and extreme self-condemnation.

Results and Discussion

The ratings on direction of aggression were found to have a moderate positive skewness when the distribution of the total sample was plotted. This skewness, however, was not considered sufficient to prevent the use of normal curve statistics. A significant biserial coefficient of correlation of .67 was obtained between the ratings and status one year after discharge from the hospital (see Table 1). Those patients in the total sample who received low ratings tended to remain unimproved after treatment. When the relationships between other significant variables and direction of aggression were studied, many showed significant intercorrelations. Table 2 summarizes these relationships. Patients within rating scale points one to three tended to have had less precipitating stress, less affective display, a longer period of illness, an insidious type of onset, and to be unmarried, disoriented, and to carry a diagnosis of schizophrenia. The diagnostic categories for those in the high ranges of the scale tended to be manic-depressive and involutional psychoses, and reactive depressions. Those earning middle ratings were most often diagnosed as having affective psychoses or psychoneuroses. The middle and high groups on the scale showed more affect, greater stress, more pronounced onset, very little disorientation, and shorter duration of illness.

Table 1

Summary Statistics for Total Group and Controlled Sample Group

Statistic	Total group	Controlled sample group
N	486	200
Mean†		
Improved	5.02	4.20
Unimproved	3.16	3.69
Standard deviation		
Improved	1.72	1.28
Unimproved	.95	1.33
t	14.31*	2.83*
r_{bis}‡	.67*	.23*

* Significant at .01 level.

† In rating scale points.

‡ Between direction of aggression and outcome.

Table 2

Relationship between Direction of Aggression and Significant Variables ($N = 486$)

Variable	Statistic	p
Affective expression	$r = .24$	.01
Orientation	$\chi^2 = 12.8$	.01
Diagnosis	$\chi^2 = 132.9$	.01
Type of onset	$r = .20$	.01
Duration of illness	$\chi^2 = 24.3$	.01
Precipitating stress	$r = .23$	.01
Treatment	$\chi^2 = 18.9$	.01
Marital status	$\chi^2 = 9.4$	$>.05$

Table 3 presents a summary of the results of matching the improved and unimproved groups on eight of the significant variables, while the aggression variable was left free to vary. The biserial coefficient of correlation was .67 before matching, and .23 when other relevant factors were held constant, as is shown in Table 1. The

Table 3

Differences between Significant Variables in the Controlled Sample Group

Variable	Improved	Unimproved	Statistic	p
Affective expression				
Mean rating	5.38	5.26	$t = .48$	$>.05$
Orientation				
Oriented (N)	91	86	$\chi^2 = 1.23$	$>.05$
Disoriented (N)	9	14		
Type of onset				
Mean rating	5.11	5.31	$t = .77$	$>.05$
Duration				
Mean, in months	32.4	28.6	$t = .96$	$>.05$
Precipitating stress				
Mean rating	3.03	2.90	$t = .52$	$>.05$
Type of treatment				
Non-physical (N)	30	35	$\chi^2 = .57$	$>.05$
Physical (N)	70	65		
Marital Status				
Single (N)	56	55	$\chi^2 = 5.23$	$>.05$
Married (N)	41	34		
Divorced	3	11		
Diagnosis (N)				
Schizophrenia	45	45	None	
Affective psychoses	11	11		
Other psychoses	5	5		
Psychoneuroses	25	25		
Other nonpsychoses	14	14		

means for the improved and unimproved groups were 4.20 (SD, 1.28) and 3.69 (SD, 1.33), respectively, with a t of 2.83. Although this difference is still significant at the 1 per cent level of confidence, it represents quite a loss in predictive power. However, the retention of a significant relationship to outcome of treatment after the rather stringent method of partialling employed here indicates the probable potency of direction of aggression as a prognostic variable.

The method used here appears to serve two functions. The first is concerned with techniques of studying a single, well-defined variable while controlling the concomitant influence of related factors. The second function is that of suggesting the type of objective measurement which may facilitate comparisons of many studies. A single example may illustrate this point. Paster and Holtzman [7] reported a good prognosis for "paranoids and catatonics," results which at first seem to contradict the present findings. However, careful reading reveals essential agreement. Their improved group consisted of "paranoid schizophrenics who were dull and apathetic" or "manic-depressives with paranoid trends" and "catatonic schizophrenics who were not difficult to manage." It is obvious that one cannot fit all paranoid reactions into one paradigm or all depressive behavior into another. The use of a scale for rating direction of behavior may help to reduce the many differences in criteria which Bellak deplores [2].

Summary

A group of 486 patients hospitalized for "nonorganic" mental illnesses were categorized as improved and unimproved one year after discharge from the hospital. Twenty-one variables, extracted from the case history material, were studied for their relationship with follow-up status. Nine pertinent variables were found to be significantly related to outcome of treatment. One of these, direction of aggression, a rated variable, was found to have a significant relationship with follow-up status when the effects of the other significant variables were partialled out. It was concluded that direction of aggression, as defined, was a reliable prognostic variable. The patients who tended to direct blame or hostility exclusively onto the environment appeared to have a poorer prognosis for status one year after discharge from a mental hospital.

Received September 8, 1953.

References

1. Albee, G. W. Patterns of aggression in psychopathology. *J. consult. Psychol.*, 1950, 14, 465–468.
2. Bellak, L. *Dementia praecox*. New York: Grune & Stratton, 1948.
3. Blair, D. Prognosis in schizophrenia. *J. ment. Sci.*, 1940, 86, 378–413.

4. Feldman, Dorothy A. Studies of prognostic criteria in case-histories of hospitalized mental patients. II. Direction of aggression. Unpublished doctor's dissertation, Univer. of Pittsburgh, 1952.

5. Kalinowsky, L. B., & Hoch, P. H. *Shock treatments.* New York: Grune & Stratton, 1946.

6. Pascal, G. R., Swensen, C. H., Feldman, Dorothy A., Cole, Mary E., & Bayard, Jean. Prognostic criteria in the case histories of hospitalized mental patients. *J. consult. Psychol.,* 1953, **17**, 163–171.

7. Paster, S., & Holtzman, S. C. A study of one thousand psychotic veterans treated with insulin and electric shock. *Amer. J. Psychiat.,* 1949, **105**, 811–814.

8. Rickles, N. K., & Polan, C. E. Causes of failure in treatment with electroshock. *Arch. Neurol. Psychiat.,* 1948, **59**, 337-346.

9. Smith, Dorothea. An analysis of aggression among patients in a mental hospital. Unpublished doctor's dissertation, Univer. of Pittsburgh, 1951.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

Prognostic Significance of Precipitating Stress in Mental Illness[1]

Mary E. Cole[2]
Psychological Service of Pittsburgh

C. H. Swensen,[2] and G. R. Pascal[2]
University of Tennessee

Precipitating factors or precipitating stress is frequently reported in the literature as having prognostic value in mental illness. The consensus is that the presence of some precipitating factor, especially one of a fairly severe nature, makes for a better prognosis in mentally ill patients than does the absence of a precipitating factor. The rationale for using precipitating stress as a prognostic indicator is that persons who succumb only to fairly severe precipitating factors have a greater stamina or resistance to ordinary life stress than do those who become ill with little or no precipitating stress. This greater resistance to severe stress implies that, when such patients are removed from the stressful situation, as in hospitalization, they have a better chance for a remission of symptoms than would be true if they had less resistance to stress.

Many studies have been made of precipitating stress, and a preceding article [6] contains an extensive bibliography. These studies can be summarized very briefly by stating that most of them agree that the presence of precipitating stress is generally indicative of a better prognosis than the absence of precipitating stress. In most studies little attention is given to statistical treatment of the data beyond reporting percentages. In none of the studies which dealt with multiple prognostic variables was any attempt made to study one of the significant variables while controlling the others. These latter two considerations are the basis for the present research.

Procedure

The case records used in this study were those of the Western Psychiatric Institute and Clinic (WPIC), University of Pittsburgh. The records of all patients admitted to this hospital since November 1942 and discharged not later than July 1950 (to allow a minimum of one-year follow-up) were examined. There were 1,118 admissions and discharges between these dates. The following criteria were then applied to these records for selective purposes: age 15 to 55, lack of known cortical damage, absence of mental deficiency, absence of incapacitating somatic illness, and absence of previous attacks of mental illness. The last criterion requires further explanation. Since WPIC is primarily a referring hospital, most of its patients come from other hospitals. Therefore, the criterion "absence of previous mental illness" means that upon admission to WPIC the patient was suffering from one continuous attack of mental illness without known prior attacks which resulted in referral to a psychiatrist. Application of these criteria left 486 cases available for further study.

Status one year after discharge was categorized as "improved" or "unimproved." A patient was considered improved if, one year after discharge from the WPIC, he was working at a level of efficiency roughly equivalent to that which obtained prior to illness. He was considered unimproved if, one year after discharge from the hospital, he was in another mental hospital, or was manifesting other behavior which, when judged by comparison with expectations about a person of like status, was indicative of inability to cope with his environment.

---

[1] Based upon a dissertation submitted to the Graduate School of the University of Pittsburgh by the senior author in partial fulfillment of the requirements for the Ph.D. degree.

[2] At the University of Pittsburgh when this research was done.

Twenty-one variables were selected for study in the sample of 486 cases. The variables were: sex, age at admission, marital status, religion, nationality, occupation, economic level, body type, education, intelligence, affective expression, orientation, direction of aggression, diagnosis, patient's age at onset of illness, type of onset, duration of illness, total hospitalization, status on discharge, precipitating stress, and treatment. A more complete treatment of the measurement of each of these variables and their relationship to outcome of illness was reported by Pascal *et al.* [6]. However, since precipitating stress is the special province of this paper, it will be dealt with here in detail.

Rating of Precipitating Stress

A nine-point scale was constructed for rating precipitating stress. Point 1 on the scale represented little or no stress, and point 9 represented severe stress. A condensation of the final scale used to rate the cases is given below. The full scale may be found in Cole's dissertation [3].

Point 1. Little or no precipitating stress.

Included here are those cases in which no precipitating factor was found. Many of these will have case histories of a lifetime of stress and poor adjustment with no period of relatively stable adjustment. In other cases, adjustment may have been fairly adequate, with the mental illness precipitated by some trivial, often temporary, kind of stress which the ordinary person frequently encounters and copes with easily.

Point 3. Relatively mild precipitating stress.

Included here are occurrences which are admittedly somewhat stressful, but which are so common to a great many people that they may be regarded as ordinary life stresses. Also, for most people, the results from this amount of stress are usually temporary in nature. General examples are: induction into military service of a young single man, death of a parent when one is an adult, birth of a child, illness in the family, and mild physical illness or minor injury to the patient.

Point 5. Moderate precipitating stress.

Included under this point are such precipitating factors as induction into military service of a man with family responsibilities, heavy financial responsibilities, moderate illness or injury to the patient or his family, poor marital adjustment, worry about a job, chronic difficulties with in-laws, and the like. Also included in Point 5 are combinations of two or more instances of milder stress, usually occurring within a relatively brief period of time so that they have the cumulative effect of moderate precipitating stress.

Point 7. Relatively severe precipitating stress.

This category of the scale includes instances of environmental stress more severe than the average person may ever experience, and usually involves the loss or threatened loss of something very important to the individual's security. It is exemplified by the threatened loss of a loved one through serious illness or injury, rejection by a loved one, or loss of livelihood after a period of good adjustment. It will also frequently include a combination of mild or moderate stresses occurring together and resulting in the equivalent of severe precipitating stress.

Point 9. Extraordinary, extreme precipitating stress.

This category includes precipitating factors so extreme as to be overwhelming, such as loss of a child or children, or of a spouse where the relationship has been a very close and dependent one. It would also include the complete disruption of a well-adjusted life pattern through some catastrophe or disaster striking suddenly, or the same effect caused by a period of severe, mounting environmental stresses following a long period of stable adjustment. In any event, a case history rated point 9 would be the net result of a far greater amount of precipitating stress than the average person would ever experience.

It will be noted that only five points (points 1, 3, 5, 7, and 9) on the scale were defined. The even-numbered points (points 2, 4, 6, and 8) were used as intermediate points for the rating of cases that appeared to fall between the defined points on the scale.

After the scale points were defined and illustrated, forty cases were selected and independently rated by two clinical psychologists for amount of precipitating stress, using the scale just described. Their ratings correlated .90, indicating fairly good reliability.

The relationship between the variables and outcome. Eleven of the variables (affective expression, orientation, direction of aggression, diagnosis, type of onset, duration, hospital stay, marital status, status on discharge, precipitating stress, and treatment) are significantly related to outcome of illness at the .01 level or better. A more complete treatment of each of these variables is contained in a previous paper [6].

Method used to control the significant variables. The crucial part of this study lies in the selection of the controlled sample population, and in the controlling of all variables significantly related to outcome except the variable of precipitating stress. Precipitating stress was permitted to vary, and appropriate statistics were then used to determine whether this variable alone was related significantly to status one year after discharge.

Since the criterion used throughout was status one year after discharge, categorized as improved or unimproved, the 486 cases were divided into an improved group of 232 and an unimproved group of 254. These two groups were then matched for the variables found to be significantly related to outcome, other than precipitating stress, by eliminating cases until there were no significant differences in means between the improved and unimproved groups with respect to each of these variables. These controls reduced the improved and unimproved groups to 90 cases each. The differences between the final groups on the significant variables are contained in Table 1.

Table 1

Significance of Differences between the Controlled Sample Groups on the Controlled Variables

Variable	Statistic	*p* value
Affective expression	$t = .14$	$>.05$
Orientation	$\chi^2 = 3.24$	$>.05$
Direction of aggression	$t = .00$	$>.05$
Type of onset	$t = .50$	$>.05$
Duration of illness	$t = .58$	$>.05$
Total hospital stay	$t = .63$	$>.05$
Treatment	$\chi^2 = .00$	$>.05$
Diagnosis	$\chi^2 = .00$	$>.05$
Marital status	$\chi^2 = .10$	$>.05$

As a further check, the final controlled sample groups were compared with respect to the other ten variables not found to meet the level of confidence of .01 decided upon. It was found that the controlled sample groups did not differ significantly in means for any of these nonsignificant variables.

Results and Discussion

When the controlled sample groups were finally matched on the significant variables, they were compared with respect to precipitating stress, which was the only factor left free to vary. It was found that there were no statistically significant differences between the improved and unimproved groups in amounts of precipitating stress. Table 2 shows the differences between the controlled sample groups in precipitating stress. A *t* of .35 was obtained, which is not statistically significant. These results were strikingly different from those obtained in the original total sample population of 486 cases, where precipitating stress was found to be related to status one year after discharge at the .01 level of confidence. It will be noted that the mean rating for the total improved group was 3.23, and for the unimproved group, 2.51. However, in the controlled sample groups, the mean rating for the improved group was 2.92, and for the unimproved group it was 3.01. The difference between the total groups was significant, but the difference between the controlled groups was not. The average amount of precipitating stress for this particular sample of hospitalized patients fell at point 3 on the scale. Point 3 includes what might best be described as ordinary life stress: those situations which admittedly are somewhat stressful, but which occur so commonly and are withstood by such a

Table 2

Summary Statistics for the Total Group and the Controlled Sample Group

Statistic	Total group	Controlled sample group
N	486	180
Mean*		
Improved	3.23	2.92
Unimproved	2.51	3.01
Standard deviation		
Improved	1.91	1.60
Unimproved	1.60	1.81
t†	4.43§	.35
r_{bis}‡	.25§	−.04

* In rating scale points.
† Between improved and unimproved groups.
‡ Between stress and outcome.
§ Significant at .01 level.

Table 3

Relationship of Precipitating Stress to Other Significant Variables

Variable	Statistic	*p* value
Affective expression	$r = .18$	.01
Orientation	$\chi^2 = 6.5$	.05
Direction of aggression	$r = .23$	.01
Diagnosis	$\chi^2 = 12.9$	.02
Type of onset	$r = .11$	.05
Duration of illness	$\chi^2 = 19.3$	.01
Treatment	$\chi^2 = .81$	.50
Marital status	$\chi^2 = 22.24$	.01

large proportion of the population that on this scale they were rated as mild.

To show the difference between the total sample and the controlled sample groups with respect to the prognostic value of the variable, precipitating stress, Table 3 is included, showing the relationships between the eight significant variables and precipitating stress. An examination of these relationships reveals that precipitating stress is significantly related, with a probability value of .01, to the following four variables: affective expression, marital status, direction of aggression, and duration of illness. Relationships approaching significance (less than .05 but greater than .01) are found with orientation, diagnosis, and type of onset. It would appear from these results that precipitating stress is not a pure variable, but rather, that it depends for what prognostic value it has on its relationships to other significant variables.

Most of the significance found between amount of affective expression and precipitating stress derived from the upper end of the precipitating stress scale. That is, patients having more than ordinary life stress as a precipitating factor tended to be rated high on affect. Those patients with very little precipitating stress varied in the amount of affective expression.

With regard to marital status, more single persons became mentally ill with very little precipitating stress than did married ones. However, more married people broke down under severe to extreme precipitating stress than did single people. This latter finding might be interpreted in the light of the additional personal and financial responsibilities that go with marriage, thereby creating more opportunities for stressful situations to occur.

The closest relationship between precipitating stress and any of the significant variables was that found with direction of aggression. Where precipitating stress was high, there was a tendency for aggression to be turned toward the self. Where precipitating factors were few, aggression was found to be more often projected outward toward the environment. Direction of aggression was also significantly related to diagnosis, which perhaps, in turn, gives the approach to statistical significance found between precipitating stress and diagnosis.

As might be expected, duration of illness was found to be significantly related to precipitating stress. In those cases of very long duration, little or no precipitating stress was found, and where the illness was of relatively brief duration, greater amounts of precipitating stress were found.

Summary

Using predetermined criteria, 486 case records were selected from the files of the Western Psychiatric Institute and Clinic. These records comprised all cases admitted up to July 1950 which met certain criteria. All were records of patients between the ages of 15 to 55 who were without mental deficiency, without known damage to the cortex, without incapacitating somatic illness, and without previous attacks of mental illness. The patients' status one year after discharge from the hospital was determined and categorized as improved or unimproved. The records were examined and 21 variables extracted which were, by appropriate statistical techniques, related to status one year after discharge. Of the 21 variables, 11 were found to be of prognostic significance at the .01 level of confidence. These were: affective expression, orientation, direction of aggression, diagnosis, type of onset, duration of illness, length of hospital stay, marital status, status on discharge, precipitating stress, and amount of treatment.

This paper is concerned with precipitating stress as a prognostic variable when all other significant variables are controlled A nine-point scale was constructed for rating precipitating stress. Appropriate reliability and validity checks of the scale were made before it was used on the total population of cases.

After all 486 cases in the total sample population were rated on the scale for precipitating stress, they were divided into two criterion groups: those improved one year after discharge from WPIC, and those who remained unimproved one year after discharge. There were 232 in the improved group and 254 in the unimproved. These two groups were then matched with respect to all the variables found to be significantly related to outcome by withdrawing cases until there were no significant differences between means and standard deviations for each variable between the improved and unimproved groups. These controlled sample groups numbered 90 improved and 90 unimproved cases. A further check on these two groups compared the means of the ten variables which did not meet the criterion of a probability value of .01, and no significant differences were found between the controlled sample groups on any of these variables. Thus, with all significant and nonsignificant variables controlled, the variable of precipitating stress was the only one left free to vary. The result was that no statistically significant differences were found between the improved and unimproved groups with respect to the amounts of precipitating stress involved in their illnesses.

Received August 11, 1953.

References

1. Bellak, L. *Dementia praecox*. New York: Grune & Stratton, 1948.
2. Cheney, C. W., & Clow, H. E. Prognostic factors in insulin shock therapy. *Amer. J. Psychiat.*, 1941, **97**, 1029–1039.
3. Cole, Mary E. Studies of prognostic criteria in the case histories of hospitalized mental patients: III. Precipitating stress. Unpublished doctor's dissertation, Univer. of Pittsburgh, 1952.
4. Hamilton, D. M., & Wall, J. H. The hospital treatment of dementia praecox. *Amer. J. Psychiat.*, 1948, **105**, 346-352.
5. Hunt, R. C. Relation between precipitating situation and outcome in manic-depressive psychosis. *Amer. J. Psychiat.*, 1938, **95**, 65-72.
6. Pascal, G. R., Swensen, C. H., Feldman, Dorothy A., Cole, Mary E., & Bayard, Jean. Prognostic criteria in the case histories of hospitalized mental patients. *J. consult. Psychol.*, 1953, **17**, 163-171.
7. Strecker, E. A., & Willey, G. F. Prognosis in schizophrenia. *J. ment. Sci.*, 1927, **73**, 9-39.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

Patient Cancellations Following a Sudden Cancellation by the Therapist

Miriam Crowley McCue

VA Mental Hygiene Clinic, Lowell, Massachusetts

Psychotherapists and case workers often express the opinion that their canceling of a scheduled appointment will be followed by the patient's cancellation of the next appointment. It is not unusual to hear that a patient canceled, or failed to keep, an appointment "in retaliation" following a cancellation by the therapist. Most therapists will agree that the meaning of the therapist's cancellation depends on the patient and upon his relationship with the therapist at the time. Some patients may be anxious to see the therapist after his absence, and their reasons for this are varied. Others may consciously or unconsciously feel that "if he can cancel, I can too." Still others may think that the therapist is not interested in them, and they may stay away with hurt feelings, or stay away because their own interest begins to lag after an interval without treatment.

Resistance to treatment may follow the therapist's absence, and it may appear in various forms of behavior. One of the most effective modes of resistance is canceling the appointment, for this precludes any interaction between the therapist and patient. If a patient resents his therapist's absence, yet keeps his appointments regularly and on time, the therapist may note resistance in the manner of the patient or in the content of the interview material. The more the patient conceals his resentment, the more difficult is the task of recognizing overt and covert signs of hostility.

Since many varieties of response are possible for the patient, several avenues are open for the study of what sudden cancellations mean to him. It seems in order to start with a simple objective approach and determine first just how often patients cancel appointments after their therapists have canceled appointments with them.

If the therapist is uneasy about canceling, such a count will tell how valid is his concern. Perhaps therapists should ask themselves why they believe that patients will cancel after therapists do. If the likelihood of a patient's acting-out is great, the therapist's anxiety may increase — which only complicates treatment.

The present study was undertaken to discover whether or not patients actually do cancel appointments after their therapist cancels, as many therapists seem to believe.

Method

Subjects. The data were obtained from the attendance cards of 94 male veterans of World War II who were currently or previously in treatment at a mental hygiene clinic in a large mill city in New England. There was little variation in the age, education, and occupational level of the patients. Most of them had weekly appointments, although some were seen twice a week, and some once or twice a month. Therapists of these patients were seven staff members of the clinic, specifically two psychiatrists, four social workers, and one psychologist.

Procedure. The appointment records were examined for instances in which a therapist canceled an appointment with a patient. Only unexpected cancellations were considered, i.e., those of which patients did not know until the day of the appointment. The first such cancellation was the only one counted for any one patient. The experimental group, *E*, made up of 47 *S*s, was drawn from these cases.

A control group, *C*, consisted of 47 *S*s who *kept* the appointment which matched the therapist-canceled appointment of one of the *E* group members. Thus 47 pairs of *S*s were matched for the number of the appointment from the start of therapy; this shall be referred to as the "matched appointment." Matched appointments ranged from the fifth one after therapy started to the seventy-eighth, the median being the sixteenth. The two groups were further matched patient for patient, each pair having the same therapist, approximately the same total number of appointments, and treatment dates at similar seasons wherever possible. No one was selected for a control who had ever had an appointment canceled by his therapist up to, and including, the sessions investigated here.

All *S*s had at least four appointments before, and four after, the matched appointment. The attendance or absence of each *S* was recorded for each of these eight appointments.

Results

Table 1 presents the number of cancellations by both groups on the eight appointments encompassing the matched appointment. The term cancellation is used to include any absence from an appointment, whether *S* actually canceled it, or merely failed to keep it.

Table 1

Number of Patient Cancellations before and after the Matched Appointment

Appointment	Group *E*	Group *C*
Before		
4th	6	11
3rd	14	16
2nd	7	11
1st	9	17
Matched appointment	—	0
After		
1st	6	17
2nd	11	12
3rd	13	10
4th	10	10

The data in Table 1 do not support the hypothesis that a patient will cancel his next appointment after his therapist cancels. There is, in fact, a decrease in the number of patient cancellations on that appointment. Only six *S*s in group *E* canceled the appointment which followed the appointment canceled by their therapists. This figure is only 14.9 per cent of the 47 cases of therapist cancellations. In other words, about 85 per cent of the patients attended the first appointment following the one that the therapist suddenly canceled. Patients seem to be more anxious to see their therapists after missing an appointment than to try to "get even" by canceling. The degree of association between therapist and patient cancellations and attendances as determined by the phi coefficient is —.88. This value indicates that there is a marked negative association between these variables.

Confirmation of one aspect of the negative relationship discovered here may be found in a comparison of the number of cancellations made by each of the two groups on the first appointment after the matched appointment. Cancellation by the therapists is followed by high patient attendance, and attendance by all the therapists is followed by high patient cancellation. The chi-square test was applied to find the significance of these data. A fourfold contingency table was constructed showing the association of patient attendance and cancellation in the two groups. The obtained chi square of 6.96 exceeds the .01 level of confidence for 1 *df*. The two groups can therefore be said to differ significantly with respect to these cancellations. There were fewer cancellations on the first appointment after the matched appointment in group *E* than in group *C*, and this difference is not the result of sampling variations.

Discussion

The cancellation frequencies in this study suggest trends which could be validated only by the selection of new groups. Perhaps the greater frequency of "before" cancellations in group *C* may be explained by the fact that the *S*s for group *C* were selected because they had kept a certain appointment. This selection may have increased the chances of getting *S*s with a high number of cancellations just prior to the matched appointment. Cancellations by *S*s in group *C* are consistently higher than those by *S*s in group *E* on the four "before" appointments, and are strikingly so on the first appointment before the matched one. Where 100 per cent attendance occurs, a fairly high amount of cancellation has gone before. Whether canceling is done by the therapist or by the patient may not be important—at least for the attendance record.

Matched pairs of Ss were used to determine whether or not the two groups were selected from the same population, and did not differ with respect to their frequency of cancellation on the four appointments preceding the matched one. Subtracting the number of "before" cancellations by each S in group C from those of the paired S in group E yielded an M_{dif} of —.40 with an SE_{dif} of .19. A t value of 2.16 was found, and with 47 df, p is between .02 and .05, indicating that this difference must be regarded as one of doubtful significance.

Testing the groups for homogeneity in terms of their frequency of cancellations over the whole period studied, however, indicates that the groups were drawn from a single population. With matched pairs of Ss, difference scores were obtained by subtracting the number of cancellations in group C from the number in group E for the four appointments before the matched appointment and also for the four appointments after it. The difference between the number of "before" and "after" cancellations was then found for each pair. The M_{dif} was .13 with SE_{dif} .24. The t value of .54 indicates that the M_{dif} is not significant ($p = .60$, $df = 47$).

Summary

The purpose of this study was to investigate the hypothesis that patients will cancel the appointment which follows an appointment suddenly canceled by their therapists. Appointment records were used to obtain a count of the cancellations of patients for the four appointments before and four after the matched appointment selected for observation. The Ss were 47 clinic patients whose therapists had suddenly canceled an appointment with them, and 47 patients whose therapists had kept an appointment which they (the patients) also kept. The records revealed that only six of the 47 Ss whose therapists were absent for an appointment canceled their next appointment. Thus the main hypothesis was not verified. A statistical analysis of the data revealed a significant difference between the groups with respect to the number of cancellations on the appointment immediately following the matched appointment. Incidental findings support the possibility that high cancellation frequency is followed by high attendance. In the control group, the highest number of cancellations occurred not only immediately after, but also immediately before, the matched, i.e., kept, appointment. Further studies on new samples are needed to confirm discoveries made in this tally.

Received August 10, 1953.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

The Comparative Effectiveness of Individual Psychotherapy, Group Psychotherapy, and a Combination of These Methods[1]

George O. Baehr
VA Hospital, Danville, Illinois

A large body of psychotherapists [8] regards with skepticism the notion that group methods can be effective in the treatment of emotional disorders. Yet, psychotherapists who have applied group methods systematically, notably Slavson [9], Klapman [5], and Dreikurs [2], have contended that group psychotherapy is not only effective but that it affords certain benefits not obtainable through individual therapy. According to this contention, individual and group psychotherapy can be expected to supplement one another to the greater benefit of the patient. The present investigation was designed to test the hypothesis that, within the limits to be set forth later, treating patients by a combination of group and individual psychotherapy is more effective than treating them by one method alone.

Procedure

Effectiveness, as a measure of therapy, does not have a universally accepted criterion. Effectiveness of therapy might be evaluated in terms of the patient's tendency to act out antisocial impulses, it might be evaluated with respect to his ability to plan realistically for the future, or again, it might be evaluated in terms of the patient's ability to hold a job and earn a living, etc. The criterion of therapeutic effectiveness employed in the present study is *discontentment*. Discontentment was chosen as the criterion because it is the central problem of the patient who voluntarily applies for treatment, because it is regarded by the investigator as the focal expression in consciousness of all intra- and interpersonal conflict and malfunction, and because it can be represented in quantitative terms. Conceptually defined, discontentment is the sense of ill-being coupled with dissatisfaction. Operationally, it is defined by a scale which was especially designed to measure the subject's discontentment level. This scale, as viewed by the subject, allows him "to show how much he is bothered" by each of his symptoms.

The Discontentment Scale [1, pp. 44-57] consists of 230 statements describing potentially discontentment-arousing experiences involving persons, the supernatural, sex, body functions, thought processes, the senses, the ego ideal, the superego, etc. The subject is required to place each statement which is true about him on an eight-interval scale to indicate the intensity with which each experience troubled him during the past week.

The average value that a subject assigns to the statements is taken as the best single index of his level of discontentment during the week immediately preceding the examination. It needs to be emphasized that the Discontentment Scale yields an index of the intensity of experience. The index is not based on the number of experiences, as in the MMPI [3], nor is it based on the frequency with which experiences occur, as in the second section of the Inventory of Psychological Problems developed by the Air Force during World War II [4, p. 71].

By means of a suitable formula, the scale values are converted to discontentment indices ranging from 0 to 100. For a sample of 86 newly hospitalized patients, the 24-hour test-retest reliability of the discontentment index was found to be +.84; whereas for a sample of 28 patients about to be discharged after the completion of their treatment, the 6-hour test-retest reliability was found to be +.96. On two samples ($N = 10$; $N = 11$), the discontentment index was found to have an average product-moment correlation of +.47 with the ward physicians' estimates of the patients' need for treatment.

---

[1] From the Veterans Administration Hospital, Danville, Illinois. The data for this study were collected at the Veterans Administration Hospital, Hines, Illinois.

Each subject was given the Discontentment Scale at the beginning and end of his period of hospitalization. The effectiveness of psychotherapy for each individual was determined by calculating his relative movement index using the formula:

$$m_r = \frac{100\,(d_i - d_t)}{d_i}$$

where

m_r is the relative movement in the discontentment level,

d_i is the discontentment level before therapy (the initial discontentment level), and

d_t is the discontentment level after therapy (the terminal discontentment level).

The relative movement index has been found to have a product-moment correlation of +.66 ($N = 66$) with the patients' own evaluation of the amount of benefit derived from their period of hospitalization.

In the present investigation, the experimental population consisted of 66 World War II hospitalized, male veterans ranging in age from 20 to 58 years, with the median age at 30. They met the following criteria: voluntary admission for treatment, voluntary discontinuance of treatment and discharge from the hospital, a diagnosis of psychopathology, and legal competence. These criteria, rather than narrow diagnostic categories, were made the basis of selection because of the well-known unreliability of the latter [4, p. 85] and because the criterion of psychotherapeutic effectiveness employed was felt to be appropriate only for voluntary patients. Diagnostically, the experimental population consisted of 2 early schizophrenics, 1 manic-depressive of the depressive type, 40 psychoneurotics, and 23 patients suffering from character and behavior disorders. The subjects were treated with nondirectively oriented individual and group psychotherapy in various proportions and divided into the following categories:

A. Those treated predominantly by group therapy

B. Those treated predominantly by individual therapy

C. Those treated by an approximately equal amount of group and individual psychotherapy

D. Those treated predominantly by one method of psychotherapy, i.e., categories *A* and *B* combined.

Division of the total experimental population into these categories was accomplished with the aid of the formula:

$$R = \frac{G - I}{G + I}$$

where

G represents the hours of group therapy,

I represents the hours of individual therapy, and

R represents the ratio of the difference between group and individual therapy hours to the total number of hours in psychotherapy.

R has a negative value if individual therapy exceeds group therapy, has a positive value if group therapy exceeds individual therapy, and is equal to zero if the amount of time spent in each is the same.

Two different procedures were used to test the hypothesis.

1. In procedure 1 the 66 experimental patients were divided into three categories. The 17 patients (approximately 25%) who had a predominance of group therapy ($R \approx +1$) formed category *A*; the 16 patients (approximately 25%) who had a predominance of individual therapy ($R \approx -1$) formed category *B*; and the 33 patients (50%) who had an approximately equal amount of group and individual therapy ($R \approx 0$) formed category *C*. Each subject's relative movement index was calculated separately and the mean of these measures was computed for each category. These means indicate the effectiveness of the particular methods of therapy applied to the categories. The t test was employed to determine the significance of the difference between the means. Since the purpose of the investigation was to determine the functional relationship between the methods of therapy and movement in the discontentment level, precautions had to be taken to ensure that the categories were not significantly different with respect to other variables which might influence movement in the discontentment level. These variables are the number of therapy hours, the duration of hospitalization, and the initial discontentment level. In no instance did the categories differ on these variables at even the .10 level of confidence.

2. In procedure 2 the 66 experimental pa-

tients were divided into two categories. Category D consisted of those 33 patients whose R value indicated a predominance of a single method of therapy ($|R| \approx 1$); and category C, as in procedure 1, consisted of those 33 patients who had an approximately equal amount of group and individual therapy ($|R| \approx 0$). In a preliminary check it was found that these two categories differed, to a statistically significant degree, with respect to the mean number of therapy hours. In order to equate the two categories, the three patients in category D who had the smallest number of therapy hours were eliminated and the three patients in category C who had the largest number of therapy hours were eliminated [6, p. 448]. The two categories were then comparable on all control variables (cf. Table 2).

Table 1

Comparison of Group Psychotherapy, Individual Psychotherapy, and the Combination of These Methods

(The relative movement index [m_r] is the measure of therapeutic effectiveness.)

	Category (by R value)				
Item	A Predominance of group psychotherapy		B Predominance of individual psychotherapy		C Individual and group therapy approximately equal
N	17		16		33
Mean m_r	+5.35		+10.06		+23.54
t for difference between means		0.419		1.652	
			1.969*		

* Significant at the .10 level of confidence.

Table 2

Comparison of the Combination Method of Group and Individual Psychotherapy with the Situation in Which One Method of Psychotherapy Predominates

(The relative movement index is the measure of therapeutic effectiveness. N = 30 for each category.)

	Category (by |R| value)					
	C Group and individual psychotherapy		D One method of therapy predominating		Diff. betw.	
Variable	Mean	SD	Mean	SD	means	t
m$_r$ Relative movement index	+25.93	25.22	+ 6.87	31.82	19.06	2.529*
G+I Hours of psychotherapy	23.63	13.82	20.27	11.81	3.36	.997
H Duration of hospitalization in days	61.87	28.86	57.73	17.29	4.14	.662
d$_i$ Initial discontentment index	53.17	17.38	52.03	17.20	1.14	.250
d$_t$ Terminal discontentment index	37.83	15.35	45.10	14.14	7.27	1.875†

* Significant at the .02 level of confidence.

† Significant at the .10 level of confidence.

Results

Procedure 1, which divided the experimental population into three categories, led to the finding that the category treated with both group and individual therapy had the highest relative movement index whereas the category treated predominantly with group therapy had the lowest movement index. These findings are in keeping with the hypothesis of the investigation and suggest that the methods rank with respect to therapeutic effectiveness as follows: group-plus-individual therapy, individual therapy alone, and group therapy alone (the least effective). However, the findings are at best only suggestive since the only t that was significant even at the .10 level of confidence was the one between the combination method and the group method. The results for procedure 1 are shown in Table 1.

The results from procedure 2, which divided the experimental population into two categories, are given in Table 2. It will be seen from the table that there is a significant difference (at the .02 level of confidence) between the two categories with respect to the relative movement index. This indicates that the group-plus-individual method was found to be more effective than one method of psychotherapy and thus confirms the hypothesis of the investigation.

Table 2 also gives the statistical data on several other variables. Within the limits of the experimental design, three of these variables, viz., the total number of therapy hours ($G + I$), the duration of hospitalization (H), and the initial discontentment level (d_i) could inadvertently have functioned as independent variables, i.e., could have functioned to contribute to the difference between the relative movement indices. The statistical data in Table 2 indicate that this possibility is unlikely, since there are no significant differences between the two categories on these variables.

In addition to the significant difference between the two categories with respect to the relative movement index, the table also shows a difference (significant at the .10 level) in the terminal discontentment index, the category treated by the combination method having the lower terminal index. The finding of this relationship is consistent with the fact that, while the initial discontentment levels of the two categories were equal, the relative movement in the discontentment level of the category treated by the combination method of therapy was significantly greater.

Discussion

It is important that the limitations of this investigation be made explicit.

1. A completely adequate evaluation of therapeutic effectiveness would take into account not only the degree and rapidity with which therapeutic changes are brought about but also the persistence of these changes after termination of therapy. The present study deals only with the degree of change over the course of therapy.

2. The problem of the therapeutic goal, i.e., the criterion for therapeutic effectiveness, has many solutions. This research employed discontentment and the relative movement in the patient's discontentment level as the criterion and the measure of the effectiveness. [Cf. 1, pp. 9-43, for a detailed discussion on the use of this criterion.]

3. The findings are limited to voluntary patients with psychopathology. They should not be generalized, without further evidence, to involuntary patients.

4. Although secondary sources of training and experience, and differences in the personalities of the psychotherapists, are likely to have introduced a certain amount of eclecticism in therapeutic approach, the therapeutic atmosphere was, nevertheless, fairly uniform from patient to patient. The 12 persons who administered the psychotherapy were all clinical psychologists trained in the nondirective methods of Rogers [7]. The therapy emphasized personal problems and emotional expression. In addition, in group therapy, group interaction was fostered. Lectures on psychology and other topics, and the giving of deep interpretations and advice were judiciously avoided. The findings of this investigation must be understood in the light of this setting.

5. In the present research, the expression "combined method of therapy" has meant group therapy followed by individual therapy. Without further evidence, we cannot generalize to courses of psychotherapy in which this sequence is reversed or where group and individual psychotherapy are carried on simultaneously.

Within these limitations, the findings of the present investigation permit the inference that the combination of group and individual psychotherapy is more effective than one method alone. They also suggest that, with respect to their effectiveness, the three methods rank, from most to least effective, as follows: the combination of methods, individual psychotherapy alone, group psychotherapy alone.

Various writers who have had experience with both group and individual psychotherapy have recorded clinical observations that have implications for the relative effectiveness of these methods used singly and in combination. At the International Congress on Psychiatry held in Paris in 1950, Dreikurs [2, p. 5] stated that, in 1929, he quite accidentally stumbled upon the discovery that treating patients in groups affords therapeutic benefit not yielded by individual therapy alone. Since that time, he has used group therapy as an adjunct of individual therapy, clearly implying that he considers the combination to be more beneficial, criteria not stated, than either method alone.

Klapman implies a similar ranking for effectiveness when he writes, "as catalyst, furthermore, group therapy will be found efficacious in speeding the dissolution of resistances. When in individual interview the patient ceases to produce any material, 'shuttling' him into a group session may not only help in overcoming resistance but even be of diagnostic importance" [5, p. 89]. Nolan D. C. Lewis, in his foreword to a volume edited by Slavson, expressed himself more cautiously, writing, "Experts in Group Therapy do not claim that it is a substitute for individual treatment, but feel that it is indicated in special, carefully chosen cases, with a proper assignment as far as the composition of the group is concerned" [9, p. 10].

Although the experimental results from procedure 1 are not conclusive in this respect, they are consistent with the clinical observations of these writers, insofar as the methods were found to rank in the order: combined therapy, individual therapy alone, group therapy alone (least effective).

The experimental results from procedure 2 strongly suggest that the combination method of therapy is more effective than either the group or the individual method alone. A superficial explanation may be inferred from the quotations. If there are benefits to be derived from group therapy that cannot be provided by individual therapy under certain circumstances and if the converse is also true, then, barring the possibility of interferences, we would expect their combination to produce greater benefits than either alone.

Although the present study and Dreikurs' observation come to the same conclusions regarding the relative effectiveness of the three methods of therapy considered, our experimental conditions provided an essentially nondirective type of therapy (client-centered psychotherapy), while Dreikurs used a directive type (individual psychology). Also, Dreikurs reversed the sequence of the therapies, applying first individual, then group psychotherapy. The evidence suggests that there is something about combining group and individual therapy which makes for greater effectiveness, independent of the order of application and the type of therapy used. How far the type of therapy can differ from those mentioned here and still conform with this generalization is a problem for further research. Accepting the high probability that group therapy has the capacity to augment individual therapy, the further question is raised for investigation: What are the particular therapeutic ingredients that each therapy supplies?

Summary

The investigation was designed to test the hypothesis that, within the limits described, treating voluntary patients by a combination of group and individual psychotherapy is more effective than treating them by one method alone. The findings strongly suggest that while all three methods produce beneficial results, the combined method is superior to one method alone. They also suggest that the methods rank themselves, from most to least effective, in the following order: group-plus-individual psychotherapy, individual therapy, group therapy.

Received October 15, 1953.

References

1. Baehr, G. O. The evaluation of psychotherapeutic results with special reference to self-referred patients. Unpublished doctor's dissertation, Univer. of Chicago, 1951.
2. Dreikurs, R. R. Psychotherapie de groupe. Paper read at the International Congress on Psychiatry, Paris, 1950. Distributed in mimeographed form.
3. Hathaway, S. R., & McKinley, J. C. *Manual for the Minnesota Multiphasic Personality Inventory.* New York: Psychological Corp., 19[illegible].
4. Heathers, G. L. Personality and adjustment studies. In S. W. Bijou (Ed.), *The psychological program in AAF convalescent hospitals.* Washington: U. S. Government Printing Office, 1947. Pp. 59-85.
5. Klapman, J. W. *Group psychotherapy, theory and practice.* New York: Grune & Stratton, 1946.
6. Peters, C. C., & Van Voorhis, W. R. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill, 1940.
7. Rogers, C. R. *Counseling and psychotherapy.* Boston: Houghton Mifflin, 1942.
8. Shaskan, D. A. Must individual and group psychotherapy be opposed? *Amer. J. Orthopsychiat.,* 1947, **17**, 290-292.
9. Slavson, S. R. (Ed.) *The practice of group therapy.* New York: International Universities Press, 1947.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

Body Size and Body-Cathexis

Sidney M. Jourard and Paul F. Secord
Emory University

In an earlier paper [1], the writers described a method for the measurement of body-cathexis, and reported that body-cathexis — the degree and direction of feelings toward one's own body — is related to anxiety, insecurity, and feelings toward the self. The present study tests whether the actual measured size of body characteristics is also a correlate of body-cathexis.

This expectation arose from consideration of the popular stereotype of the male ideal, exemplified by the tall, broad-shouldered, muscular hero in fiction and the movies.

Method

Subjects. Sixty-two male undergraduates from the introductory psychology course at Emory University served as *S*s.

Procedure. Each *S* filled in a body-cathexis questionnaire. This form, described in detail elsewhere [1], but since modified slightly, consists of a listing of 40 aspects of the body. The respondent indicated the nature of his feelings toward each item in accordance with the following five-point scale:

1. Have strong positive feelings.
2. Have moderate positive feelings.
3. Have no feeling one way or the other.
4. Have moderate negative feelings.
5. Have strong negative feelings.

Only a few *actual size* measurements related to the "masculine-ideal type" were feasible. After completion of the cathexis scale, and without prior knowledge, measurements were taken from each *S* (using a tape measure and scales) of his height, weight, shoulder width, chest circumference in relaxed position, expanded chest circumference, biceps circumference with the arm hanging loosely, and biceps when contracted.

Correlations were computed between these measurements and the corresponding cathexis ratings on height, weight, shoulders, chest, and muscular strength.

Table 1
Correlations between Body Measurements and Body-Cathexis

Body-cathexis	Body measurements	r*
Height	Height	—.55
Weight	Weight	—.04†
Shoulders	Width of shoulders	—.36
Chest	Chest (relaxed)	—.33
Chest	Chest (expanded)	—.37
Muscular strength	Biceps (relaxed)	—.52
Muscular strength	Biceps (contracted)	—.54

* $r = .33$ at .01 level for $df = 60$.

† eta for weight and cathexis = .19; eta for $\frac{\text{weight}}{\text{height}}$ and weight-cathexis = .34 ($F = 1.09$, $p > .05$).

Results

As may be seen in Table 1, correlations significant beyond the .01 level were obtained for all variables except weight. The *r*'s signify that a significant portion of the variance in cathexis scores is accounted for by variation in actual size. The direction of the *r*'s indicates that large size is associated with strong, positive feelings toward the respective body part, and small size with weak or negative feelings.

Discussion

With the exception of body weight, the findings reveal that cathexis for body characteristics pertaining to "masculinity" is related to the size of relevant body parts. Large size is apparently a desired quality among males, and its presence or absence leads to contrasting feelings toward related aspects of the male body.

The absence of a significant relationship between weight and cathexis (neither r nor η reached significance) indicates that there is no simple relationship between these two variables. The η between weight/height ratio and weight-cathexis was higher but still insignificant.

Received September 24, 1953.

Reference

1. Secord, P. F., & Jourard, S. M. The appraisal of body-cathexis: body-cathexis and the self. *J. consult. Psychol.*, 1953, **17**, 343-347.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

Validation of Rorschach Criteria for Predicting Duration of Therapy[1]

Robert G. Gibby, Bernard A. Stotsky
VA Mental Hygiene Clinic, Detroit, Michigan
E. Wesley Hiler, and Daniel R. Miller
University of Michigan

Within the past two years, there have appeared in this journal four articles that dealt with the use of the Rorschach test to predict duration of therapy. Although they used similar techniques and populations, these studies contradicted each other in a number of major aspects. The present investigation was designed to treat some of these areas of contradiction with a view to discovering the principle which accounted for the discrepancy. In order to elucidate the present design, we will describe the primary issues in terms of the findings of the previous investigators.

Possibility of prediction. The first article, by Rogers, Knauss, and Hammond [5], reported no significant difference on 99 Rorschach signs between patients at a Veterans Administration Mental Hygiene Clinic who stayed in therapy for four or less sessions and those who remained five times or more. However, Kotkov and Meadow [4], studying veterans engaged in group psychotherapy, obtained by the use of discriminant function a weighted combination of three scores, *R*, *D%*, and *FC — CF*, which discriminated with 69 per cent accuracy between patients with fewer than nine sessions and those who stayed in nine sessions or more. In validating their formula on a second sample of patients in individual therapy, Kotkov and Meadow again found that they could discriminate with 69 per cent accuracy.

Stimulated by the success of Kotkov and Meadow, Auld and Eron [1] tried the same formula on a group of 33 patients in a general hospital with no success. The accuracy of discrimination between patients who stayed in and those who left before nine sessions was 52 per cent, which is only 2 per cent from chance. The number of responses alone discriminated significantly. With their cutting point of 19, the accuracy was 67 per cent.

The present writers did an initial study shortly after the negative results of Rogers, Knauss, and Hammond appeared. In an attempt to determine whether the Rorschach test could predict duration of therapy for extreme groups, we attempted to discriminate between those who stayed in therapy for five sessions or less and those who remained for twenty or more. We did not attempt to establish a formula, but were interested solely in the specific Rorschach signs which *discriminated* between these two extremes. We were encouraged by the finding that differences between those who terminated and those who remained were very significant for a large series of scores, although a fair proportion of them were correlated significantly with the number of responses. Because we were partially successful in that endeavor, we are following it up with the present study.

Sample and Method

Subjects were selected from among patients at the Veterans Administration Mental Hygiene Clinic at Detroit, Michigan. The 269 patients in the study were predominantly neurotics, with some character disorders and ambulatory psychotics included. The socioeconomic level was predominantly lower and lower middle. The average amount of education was eleven years; the average age, 30. Most of the patients were white, only 10 per cent being Negro. Excluded from both groups were disoriented psychotics, organics, females, and patients who were forced to terminate for reasons beyond their control, such as shift in residence, etc. The precedent established by Kotkov and Meadow was followed. By the method of discrimination function, a formula that maximally differentiated between the 33 continuing patients and the 51 terminators was derived. Then the obtained formula was validated with a sample of 75 continuing subjects and 110 terminators.

---

[1] From the VA Regional Office, Detroit, Michigan.

Cutting point. In our previous article we raised the possibility that the negative results of Rogers, Knauss, and Hammond might be attributable to their cutting point of five therapeutic sessions. It did not seem reasonable to us to assume that a person who stays in therapy for six or seven interviews can be regarded as having gained significantly more from therapy than a person who stays for only four sessions. Kotkov and Meadow, who were partially successful in making predictions, used a cutting point of nine sessions; so did Auld and Eron who, although they did not discriminate significantly in terms of Kotkov and Meadow's formula, did succeed with R alone. Some of our most significant results were based on the cutting points of five and twenty sessions.

We decided to include in the continuing group only those patients who had been in therapy at least six months and, preferably, were still in treatment at the time that the data were collected. Inclusion in the terminating group, then, required the patient to have been in therapy less than six months. The cutting point derived from this criterion turned out to be 19 sessions or less.

Although 19 sessions may seem high as a point of division for the two groups, it was felt that, in attempting to predict continuation in therapy, it would be necessary to account for every individual who left treatment against the therapist's advice. This made it necessary that the sample include an intermediate group of patients who terminated therapy between their sixth and nineteenth sessions. Presumably, this group was more motivated than were the patients who dropped out between the first and fifth sessions. In theory, one might expect this group of patients to resemble the continuing group more than did the group which dropped out of treatment early. Analysis of the Rorschach data showed this not to be true. The patients in this group were more similar to those in the one-to-five session group than to patients continuing in therapy. For instance, the average number of Rorschach responses for this group was approximately 20; for the early "drop" group, also 20; for the "stay" group, 31. In view of the test performance of the intermediate group, our decision to place them with the terminators appears to have been a wise one. In our previous article we suggested that choice of a cutting point could influence the results considerably. For instance, by choosing five as a cutting point, Rogers, Knauss, and Hammond may well have misclassified a considerable number of "drop" patients. The same issue arises with the samples of Kotkov and Meadow and those of Auld and Eron. If discrepancies should arise between the findings of the present study and those of the previous ones, one possible explanation could be the choice of different cutting points.

Results

1. *The prediction formula.* On the first sample, the formula obtained by the method of discriminant analysis that best differentiated between terminating and remaining patients was

$$Y = .00049\,R + .00212\ \text{Total}\,K + .00010\,m.$$

The term Total K refers to the sum of all the K responses, none of which was weighted in any way. We applied the formula to our first sample. By using a cutting point of .02, we were able to differentiate between those who terminated and those who remained with an accuracy of 68 per cent. On applying the formula to the second sample, we were able to

Table 1

Frequency Distribution of Y Scores for Remaining and Terminating Patients

Y-score class intervals	Sample I R	Sample I T	Sample II R	Sample II T
.06000 – up	1	0	1	0
.05000 – .05999	2	0	2	1
.04000 – .04999	0	0	0	0
.03000 – .03999	5	1	4	2
.02000 – .02999	6	7	21	11
.01000 – .01999	14	19	35	49
.00000 – .00999	5	24	12	47
Total N	33	51	75	110

predict accurately 67 per cent of the time those who remained and those who terminated.

While 67 per cent accuracy is not sufficiently high to make this formula useful to the practicing clinician, our results nevertheless corroborate the findings of Kotkov and Meadow and of Auld and Eron that some sign could be found which would discriminate significantly between continuing therapy and terminating.

It should be noted from Table 1 that the formula is much more successful in predicting termination than continuance of therapy. Approximately 84 per cent of the terminators are correctly classified in the first group. The prediction of terminators in the second sample is approximately 87 per cent accurate. We have received a communication from Meadow in which he points out that the Kotkov and Meadow formula is only successful in making predictions for the group of patients who do not continue in psychotherapy. While it yielded more incorrect predictions for the continuing group in his study, the accuracy of locating terminators was 88 per cent. He also makes the cogent point that the negative results of Auld and Eron's experiment may have resulted from the fact that they used 21 patients for their continuing group and only 12 for their noncontinuing group. We then wrote to Dr. Auld, and he supplied us with the scores of his terminators. These discriminated correctly in 8 of the 12 cases. The 67 per cent accuracy, while not as high as that of Kotkov and Meadow or as high as ours, was still much higher than the over-all accuracy for both continuers and terminators.

It may be of some interest to remark on the performance of the intermediate group of 53 patients. Using .02 as a point of division, 83 per cent were classified correctly as terminators. This group was almost indistinguishable from the early terminators, of whom 88 per cent were classified correctly. These findings support the categorization of the intermediate group as terminators.

The Kotkov-Meadow formula was applied to the 185 cases in our validation sample. Patients with a Y score of .01132 were assigned to the "continuing" group and those with lower scores to the "discontinuing" group. Sixty-one per cent of the patients were classified correctly. This is slightly better than chance but considerably below the 69 per cent accuracy obtained in the studies of Kotkov and Meadow.

2. *Productivity* (R) *as a predictor.* A common finding in all these studies was that the total number of responses (R) was consistently associated with continuance in therapy. In our sample both Total K and m correlated over .60 with R. Therefore, we decided to test the ability of R by itself to predict continuation. By inspection we found 35 responses to be the best point of division between remainers and terminators in Sample I.

Forty-eight terminators and 21 remainers gave fewer than 35 responses; only three terminators and 12 remainers gave 35 or more. Accuracy of prediction was 71 per cent in Sample I. In the validation sample 102 terminators and 51 remainers gave fewer than 35 responses, while 8 terminators and 24 remainers gave 35 or more, for a total accuracy of 68 per cent. Considering the two samples together, 69 per cent of the patients were classified correctly by use of R alone. These findings would tend to suggest that R by itself can predict as well as any group of Rorschach variables, provided that a suitable cutting point is chosen. This cutting point, however, may vary from clinic to clinic. Auld and Eron found that an R greater than 18 predicted quite well for their sample.

The question inevitably arises as to how much better than chance were these predictions. For the entire sample, chi square based on a fourfold table was greater than 30, with 35 responses as a point of division. This value was significant beyond the .001 level. We may consequently conclude that the number of cases correctly classified was better than chance.

What would happen if one predicted termination for every case? The accuracy of prediction would be roughly 60 per cent, since that is about the proportion of terminators in the sample. This prediction is only 9 per cent poorer than the best Rorschach variable. Although the difference is significant, the increment in prediction resulting from R is not very great.

3. *Productivity and intelligence.* In order to explore the reason why number of responses discriminated between continuers and termin-

ators, Auld and Eron first correlated *R* with continuance. The obtained coefficient was +.55. However, when the verbal scale of the Wechsler-Bellevue was held constant, the correlation between Rorschach productivity and continuation dropped to +.07. In fact, the IQ scores gave better predictions of continuance than the number of Rorschach responses, with a biserial correlation of +.71, which was significant at the .05 level.

The authors concluded that the Wechsler-Bellevue verbal scale was more useful than the Rorschach for predicting continuance in therapy. Before accepting as valid such use of the verbal scale, they felt that their finding based on 23 patients should be supported by evidence from cross-validation studies.

This the present authors attempted to do. In the first sample there were 79 Wechsler verbal scales available, of which 37 were of remainers and 42 of terminators. The best cutting point was empirically determined at 100. A Verbal IQ score of 100 and up was considered a favorable sign; below 100, unfavorable. Fifty out of 79 patients were correctly classified, a 63 per cent accuracy. On the validation sample, 54 out of 88 were categorized correctly, a 61 per cent accuracy.

Inasmuch as our findings gave only qualified support to those of Auld and Eron, we decided to study the relationship between *R* and Verbal IQ in our sample. For 165 cases where there were both Wechsler verbal scales and Rorschachs available, the correlation between *R* and Verbal IQ was +.24. The biserial correlation between number of responses and continuance was + .38; between Verbal IQ and continuance, +.40. With *R* partialled out, Verbal IQ correlated +.35 with the criterion. With Verbal IQ partialled out, *R* correlated +.32 with continuance. In our sample the relationship between *R* and continuance was not entirely accounted for by the correlation with Verbal IQ since *R* remained significant with the verbal score partialled out. The multiple correlation of *R* and Verbal IQ with continuance was +.50. Our findings do not support the contention of Auld and Eron that the relationship between number of responses and continuance in therapy only demonstrates that more intelligent patients are more likely to continue.

4. *Social class as a predictor.* In a personal communication to one of the authors, Auld suggested that social class may be related to persistence in therapy. The present authors undertook to test the hypothesis that socioeconomic class was related to staying in therapy, the prediction that middle class patients would stay and lower class patients and unemployed patients would leave treatment. The sample was broken down into five classes:

1. Upper middle—low managerial, highly skilled clerical, and professional workers.
2. Lower middle — semiskilled and unskilled clerical and semiprofessional workers.
3. Upper lower — skilled manual workers.
4. Lower lower — unskilled manual workers.
5. Unemployed.

For the first sample there were 76 classifiable subjects of whom 47 were classified correctly, most of them in the terminating group. Sixty-two per cent accuracy of prediction was achieved; the chi square was nonsignificant, though positive. For the validation sample, 111 out of 177, 63 per cent, were categorized correctly. The chi square of 8.63 was significant at the .01 level. Since in both samples the trend was in a positive direction — with middle class people tending to remain in and lower class people to leave therapy — the hypothesis that class is related to continuance in therapy is supported in this study.

5. *Interaction between patients and therapists.* A factor not considered in previous investigations which might exert considerable limitation upon the predictive efficiency of the Rorschach test is the skill of the therapists and the nature of their resultant interactions with the patients. We were fortunate to be able to obtain data that not only bear upon this interaction but may account in part for some of the discrepancy in the results of other investigators.

Hiler [3], and Hiler and E. L. Kelly in a study as yet unpublished, did an analysis of patient-therapist interaction in order to determine whether therapists differed in respect to the types of patients they were able to keep in treatment. The number of Rorschach responses, which was assumed to have as components productivity and motivation for treatment, was used to characterize the patients.

Their findings are reported briefly below:

1. It was found that whether or not unproductive and poorly motivated patients remained in treatment was in part a function of the therapist. Although, for some therapists, terminators did not differ significantly from remainers on *R*, for most therapists they did.

2. Patients remaining in treatment were similar in respect to productivity on the Rorschach whether the therapist was a psychiatrist, psychologist, or social worker. The same was true for patients terminating.

3. Productivity on the Rorschach was not related to continuance in therapy with the female therapists employed in this study, but was associated with continuance for patients seen by male therapists. The female therapists were able to keep in treatment patients giving few Rorschach responses, but tended to lose proportionately more productive (high *R*) patients.

4. Therapists rated as warm and friendly were able to keep in treatment more of the poorly motivated, unproductive patients (low *R*) than were therapists rated as cold and unfriendly.

5. Therapists judged to be most competent at analytically oriented therapy lost proportionately fewer productive patients than did the less competent ones.

6. There was no significant difference between therapists rated high on passivity and those rated low on passivity with regard to the proportion of nonproductive patients remaining in treatment.

Discussion

Our study confirms the findings of previous studies that the total number of Rorschach responses (*R*) is significantly related to continuance in therapy. It was also found that *R* by itself predicted as well for both samples as did the discriminant function using *R*, Total *K*, and *m*. Both *R* and the discriminant function derived from our first sample predicted better than the Kotkov and Meadow formula which had been derived from a group of similar patients.

The findings from these studies, together with significant differences between groups on Verbal IQ and social class, suggest a number of tentative conclusions: (*a*) Patients who remained were, in general, more intelligent and more productive. (*b*) They appeared to possess the ambition and drive associated with the middle class and upwardly mobile lower class persons in American society. (*c*) They showed a greater awareness of the psychological factors responsible for their problems and, although they experienced greater anxiety, they were not incapacitated by their anxiety. (*d*) Patients who remained in treatment tended, to a greater extent than the terminators, to have both the capacities and the motivational state required to make effective use of the psychoanalytically oriented therapy offered at the clinic [2, 3].

Auld and Eron correctly stressed the importance of cross validation of any preliminary findings. This we have attempted to do by validating the findings from our first sample on a second sample more than twice as large as the original. In our study the requirement of cross validation was met since the formula worked as well in the second sample as it had in the first.

Auld and Eron also pointed out the need for validity generalization; that is, applying the data to a new and different sample with different characteristics. This step would require successive validation of the findings on a wide variety of patients in different clinics. There are a number of conditions limiting application of any such formula to different populations:

1. Since the formula applies best to terminators, it may not apply as adequately in a setting where, for some reason, a larger proportion of patients stay in therapy. Our formula, derived from a group of veterans receiving treatment from a VA facility, must be assumed, until it is proved otherwise, to generalize only to similar populations. Patients attending an outpatient clinic at a private hospital where they pay fees and presumably come entirely of their own accord cannot be equated with a group of VA patients who draw pensions for neuropsychiatric disorders unless there is definite evidence to show that such groups are equivalent. While this consideration should not necessarily prevent a formula derived from one type of sample from being applied to another, the limitations of such an application must be kept in mind.

2. It has been shown that both the warmth and competence of the therapist may have an effect on the applicability of a prediction formula. It is hard to see how the operation of such factors can do anything except decrease the predictive adequacy of any formula based on test material that does not take them into account.

3. The type of therapy employed can have a direct bearing on the ability of a formula to predict. Our study was performed at a psychoanalytically oriented clinic where therapy is seen as primarily a long-term affair. At a clinic oriented to more immediate short-term goals, the style of therapy, the

choice of patient, and the criteria for determining continuation and termination would probably be considerably different from the orientation found at the Detroit clinic. In view of these different conditions, it is difficult to see how a formula could be set up to predict equally well in all situations unless these factors were all relatively unimportant in comparison with the patient's personality.

Premature termination is, then, not entirely a function of the patient's personality and motivational state, but appears to be influenced by the attitude and competence of the therapist and the type of therapy offered at a clinic. Does so-called "depth" therapy, which involves the manipulation of complex verbal concepts, make sense to individuals who are unable to handle such concepts? Would other therapeutic approaches succeed in keeping such individuals in treatment, and succeed in promoting their social adjustment? Studies such as the one reported here are needed to investigate personality differences of remainers and terminators in other settings and among therapists using different approaches to treatment.

Suggestions for Further Work in Prediction

In an unpublished study, Hiler [3] found such personal data items as social class, education, occupation, diagnosis, and initial complaints to predict continuance in therapy somewhat better than chance. The figures for these items were only slightly below those obtained by psychological test data. Not only are such data easier to collect, but combinations of such items might yield predictions exceeding the psychological tests in accuracy. Perhaps the combining of psychological test scores with personal data and data pertaining to the therapist might yield even better accuracy. For instance, a formula using scores such as Rorschach *R*, Wechsler Verbal IQ, social class, and initial symptoms might very well yield a high enough prediction to be of practical use to the clinician. These four types of data do not by any means exhaust the alternatives available to the researcher interested in the problem of predicting continuation in therapy. Certainly further research is needed in this area.

Summary

1. By means of discriminant analysis an attempt was made using Rorschach scores to predict whether patients will continue psychotherapy. A formula, $Y = .00049\,R + .00212$ Total $K + .00010\,m$, was derived. When applied to a sample of 84 patients from which the weights had been derived, it correctly classified 68 per cent of the subjects as continuing or terminating therapy. On a validation sample of 185 subjects, 67 per cent of the patients were correctly classified. On the same sample, the Kotkov and Meadow formula correctly predicted 61 per cent of the patients.

2. Using the total number of Rorschach responses (*R*) by itself, it was possible to classify correctly 71 per cent of the subjects in the first sample and 68 per cent in the second.

3. For 165 patients the relationship between *R* and continuance was +.38. With Verbal IQ partialled out, *R* continued to correlate significantly with continuance, thus contradicting the previous finding of Auld and Eron that, with Verbal IQ partialled out, *R* correlated insignificantly with continuance. The multiple correlation of *R* and Verbal IQ with continuance was +.50.

4. Social class was found to be related to continuance in therapy, with middle class subjects tending to remain in and lower class subjects to drop out of therapy.

5. The implications of these findings were discussed, especially with respect to the relative importance of such factors as type of therapy and warmth and competence of the therapist in the prediction of continuance in treatment. It was concluded that premature termination is not entirely a function of the patient's personality and motivational state. Therefore, in predicting continuance in therapy, other important factors, such as those discussed, should be considered and given their proper weight in a prediction formula.

Received August 31, 1953.

References

1. Auld, F., & Eron, L. D. The use of Rorschach scores to predict whether patients will continue psychotherapy. *J. consult. Psychol.*, 1953, **17**, 104–109.
2. Gibby, R. G., Stotsky, B. A., Miller, D. R., & Hiler, E. W. Prediction of duration of therapy from the Rorschach test. *J. consult Psychol.*, 1953, **17**, 348–354.

3. Hiler, E. W. An investigation of psychological factors associated with premature termination of psychotherapy. Unpublished doctor's dissertation, Univer. of Michigan, 1953.
4. Kotkov, B., & Meadow, A. Rorschach criteria for predicting continuation in individual psychotherapy. *J. consult. Psychol.,* 1953, **17**, 16–20.
5. Rogers, L. S., Knauss, Joanne, & Hammond, K. R. Predicting continuation in therapy by means of the Rorschach test. *J. consult. Psychol.,* 1951, **15**, 368–371.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

A Rationally Derived MMPI Scale to Measure Dependence[1]

Leslie Navran
VA Hospital, American Lake, Washington

This paper reports the development and initial use of an MMPI scale to measure dependence. An attempt to develop the scale empirically failed because experienced clinicians could not reliably rate dependence from case history summaries. A rational derivation was substituted. Sixteen judges independently specified the MMPI items that they felt to bear on dependence. They further specified the direction of response which would be made by a hypothetical dependent person who was responding frankly. Tetrachoric correlation coefficients were computed for each pair of judges, and the resulting matrix, when factor analyzed, yielded one main factor consisting of eight judges. The 157 items they collectively agreed on were screened for discriminability by internal consistency statistical procedures on a sample of 50 neuropsychiatric patients. A second very similar sample of 50 patients was used to cross validate the discriminating items so obtained. The same procedure was repeated, using Sample II first and cross validating on Sample I. Fifty-seven items survived these tests and constitute the scale at present. Reliability is .91 for the 100 patients in the derivation sample (Kuder-Richardson formula 20).

The Dependence Scale correlates .72 with *Pt*, .60 with the *D* scale, —.53 with *K*, .53 with *Pa*, and .50 with *Sc*. To date, its use with known groups has yielded these results: (*a*) 100 NP patients score significantly higher than 200 normals from the original MMPI standardization population (mean difference of 7.6 points; $p < .001$). (*b*) These normals in turn score significantly higher than 265 Stanford University graduate students (mean difference of 3.81 points; $p < .001$). (*c*) The shape of the graduate student distribution is J curved, suggesting conformity to a cultural expectation. (*d*) Within the sample of 100 NP patients, character disorders score lowest, followed by neurotics and then schizophrenics. Differences are not significant between these groups, but nonparanoid schizophrenics score significantly higher than paranoid schizophrenics ($p < .04$; $t = 2.08$).

These results are considered promising, but not necessarily conclusive. However, the scale seems worthy of more investigation and exploratory use.

Brief Report
Received March 15, 1954.

---

[1] An extended report of this study may be obtained without charge from Leslie Navran, VA Hospital, American Lake, Wash., or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4227 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $2.00 for microfilm or $3.75 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

Comparison of the Rorschach and MMPI by Means of Factor Analysis

Harold L. Williams and James F. Lawrence
Letterman Army Hospital

Although a number of research papers have been published treating separately the MMPI and the Rorschach, few comparison studies have been made. Several investigators [1, 4, 6] have explored relationships between the group forms of the two tests. These studies have not revealed striking association between them. The situation was summed up by Ellis [5] when he asserted that the MMPI does not agree well with the results of the Rorschach and other projective tests. While failure to find correlation between these widely used clinical instruments would not be convincing evidence of lack of validity for either, there seems to be an a priori basis for expecting them to correlate; that is, clinicians who use both tests may make identical predictive statements from either source (e.g., "he shows depressive tendencies").

Several separate factor analyses of Rorschach determinants and MMPI variables have now been published [10, 11, 12]. Three studies on different populations have established fair agreement as to the factorial composition of the Rorschach. Even though a general factor of productivity accounts for a high percentage of the variance in the intercorrelation matrix, the other factors found were rather similar in all three studies. The five factors found by Williams and Lawrence [10] were tentatively labeled: (*a*) productivity, (*b*) intelligence, (*c*) movement, (*d*) lack of perceptual control, and (*e*) shading.

Wheeler, Little, and Lehner [9], Tyler [7], and Welsh [8] have agreed, essentially, in describing two major and two to three minor factors among the MMPI scales. On the basis of the first two factors, Welsh[1] has developed scales which he labeled *A* and *R*. The *A* scale represents a dimension containing heavy loadings on practically all of the complaint scales, particularly *Pt* and *Sc*. Judging from item content, one finds that the scale seems related to excessive preoccupation with the psychiatrically negative aspects of the self: rumination, worry, and pessimistic attitudes. The *R* scale, based on the second major factor, is correlated positively with *D, K,* and *Hy,* and inversely with *Ma*. Welsh suggests that the scale may relate to expressive-repressive personality traits.

Problem

The purpose of the present paper is to report the results of a factor analysis of the MMPI scales, certain Rorschach determinants, and the Wechsler-Bellevue Verbal IQ of 100 psychiatric patients selected at random from the files of the Clinical Psychology Section at Letterman Army Hospital. The 32 variables which were correlated can be found in Table 1.

The MMPI symbol *Es* (variable 30) refers to a scale recently developed by Barron [2] for the purpose of predicting response to psychotherapy. The scale, which is inversely related to *A*, held up rather well on cross validation. Furthermore, among normals it was found to be correlated with high ratings on a number of characteristics, such as resourcefulness, vitality, self-direction, tolerance, self-confidence, breadth of interest, emotional outgoingness, spontaneity, and intelligence. Low scorers were described as effeminate, submissive, and introspective. He suggested, therefore, that individual differences on the scale could be considered a function of ego strength.

Table 1 shows the intercorrelations computed from raw scores for the 32 variables. While the Pearson correlation coefficient was

---

[1] Welsh, G. S. Personal communication, April, 1953.

Table 1

Intercorrelation among 31 Rorschach and MMPI Variables and Wechsler-Bellevue Verbal IQ for 100 Psychiatric Patients

Symbol	Variable No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
W	1																																
D	2	19																															
d	3	17	80																														
M	4	37	75	32																													
FM	5	31	46	23	50																												
m	6	39	31	31	43	58																											
k	7	29	29	35	12	25	25																										
K	8	69	24	21	33	20	60	25																									
F	9	32	84	80	47	16	17	34	-02																								
Fc	10	28	25	10	26	22	32	32	10	00																							
c	11	55	48	37	25	26	21	30	24	46	33																						
C^1	12	44	19	31	10	16	18	16	11	03	13	37																					
FC	13	43	40	47	47	50	55	41	30	41	40	35	30																				
CF	14	55	21	20	40	29	10	30	33	17	37	56	46	40																			
C	15	43	09	00	13	-16	-05	26	30	12	01	45	31	09	44																		
P	16	10	08	23	03	45	24	-27	-25	08	25	-01	24	05	-09	17																	
R	17	51	92	69	65	55	40	65	22	90	41	59	28	61	49	06	05																
VbIQ	18	69	25	18	30	10	00	32	29	15	12	44	12	30	22	22	-10	51															
L	19	-06	19	25	-25	00	06	18	06	19	06	20	00	-06	-06	18	18	00	-24														
F	20	-38	19	00	-25	-13	32	-37	-06	06	-12	-21	00	06	-44	18	00	00	-32	-13													
K	21	50	06	06	18	-06	00	50	32	32	-06	00	06	-24	49	18	00	18	50	21	-26												
Hs	22	-13	25	00	00	32	06	00	-06	00	-12	-25	-06	-06	13	00	-12	06	00	-06	40	-10											
D	23	-06	00	-06	06	-12	06	00	00	06	06	-25	-37	00	-32	-37	-38	06	-06	-05	41	-04	74										
Hy	24	-06	06	06	25	-06	00	06	-06	-06	-06	-06	-13	-35	00	19	-18	00	00	17	20	15	76	69									
Pa	25	-06	00	06	00	00	19	-25	18	06	00	-06	06	00	-12	00	-18	00	-24	-32	66	-16	46	51	19								
Pa	26	-06	00	00	05	00	55	-25	00	-06	-06	-10	00	-06	-25	-25	-18	-32	00	-16	76	-15	45	50	37	80							
Pt	27	-25	18	00	05	-06	00	-25	-25	06	-06	-10	00	25	-12	-25	00	00	-24	-26	67	-30	67	71	40	66	64						
Sc	28	-38	25	06	00	-25	19	-12	00	18	06	-25	00	19	-31	-23	-06	00	-40	-25	81	-25	62	62	32	77	78	92					
Ma	29	-06	00	-31	19	06	00	-06	00	-31	06	23	25	25	00	00	19	50	-16	-32	45	-10	18	00	-10	47	36	37	48				
Es	30	40	38	38	43	-06	00	40	31	38	35	-06	00	-06	50	00	35	38	39	00	-60	35	-74	-42	-48	00	-62	-87	-82	-32			
A	31	-38	00	-25	-38	-13	00	-12	-06	00	06	06	00	49	00	-25	-25	00	-31	29	67	-28	59	65	29	68	60	95	90	41	-83		
R	32	-25	-25	-06	-12	-12	-25	-12	00	38	10	-32	-06	-37	32	-06	00	-25	00	33	00	27	20	20	34	00	22	-09	-04	-58	00	-08	

Table 2

Centroid Factors

Variable No.	Symbol	I	II	III	IV	h^2
1	*W*	56	54	22	−32	75
2	*D*	76	24	−32	39	89
3	*d*	56	33	−18	18	49
4	*M*	58	30	−18	07	46
5	*FM*	43	17	−28	07	30
6	*m*	59	10	−28	−28	51
7	*k*	42	40	26	16	43
8	*K*	42	23	32	−33	44
9	*F*	61	29	−25	48	75
10	*Fc*	37	21	07	−19	22
11	*c*	51	39	04	05	42
12	*C*¹	34	21	−08	−28	24
13	*FC*	63	18	−26	−14	52
14	*CF*	41	45	37	−11	52
15	*C*	22	27	24	−28	26
16	*P*	10	02	−31	−20	15
17	*R*	87	39	−20	34	106
18	*V*	30	38	26	02	30
19	*L*	00	16	20	30	16
20	*F*	27	−75	−23	18	72
21	*K*	19	30	45	10	34
22	*Hs*	35	−67	18	38	75
23	*D*	26	−69	32	41	81
24	*Hy*	24	−48	48	43	70
25	*Pd*	42	−69	17	−35	80
26	*Pa*	32	−72	−05	−18	66
27	*Pt*	37	−85	−18	18	92
28	*Sc*	39	−87	−25	−07	98
29	*Ma*	31	−32	−31	−45	50
30	*Es*	09	88	15	−21	85
31	*A*	37	−82	14	00	83
32	*R*	−16	−19	−50	10	30

used between MMPI scales, tetrachoric coefficients were used between MMPI and other variables.[2]

With few exceptions, the Rorschach determinants are not highly correlated with MMPI clinical scales. Several have significant negative correlations, and these same variables are positively associated with *Es*. *W*, *d*, *M*, *k*, *Fc*, *CF*, and *P* average about .40 with this variable. It will be noted that *Es* has a respectable correlation with Verbal IQ, a finding which supports Barron's observations. As had been demonstrated in earlier studies, the MMPI scales are markedly intercorrelated.

The intercorrelation matrix was analyzed by Thurstone's centroid method, and the analysis was continued until five factors had been extracted, at which time the residuals varied from —.18 to +.20.[3]

The resulting factorial patterns were rotated orthogonally and obliquely, an oblique solution being obtained by the trial vector method [3, p. 204]. Unrotated and rotated factors (oblique and orthogonal) are shown in Tables 2, 3, and 4. Descriptions are based on consideration of all three solutions.

Factor I contains high loadings on most of the Rorschach determinants and does not include MMPI scales. It can best be considered

Table 3

Orthogonal Factors

Variable No.	Symbol	I′	II′	III′	IV′
1	*W*	46	70	−10	09
2	*D*	90	−15	−06	−02
3	*d*	66	07	−28	−06
4	*M*	61	13	−26	06
5	*FM*	45	−02	−33	−03
6	*m*	40	16	−47	20
7	*k*	53	33	−03	−01
8	*K*	23	55	−16	27
9	*F*	83	−17	20	−13
10	*Fc*	28	32	−17	12
11	*c*	56	29	01	−01
12	*C*¹	23	30	−09	07
13	*FC*	51	14	−65	14
14	*CF*	40	59	00	−07
15	*C*	12	49	−20	−10
16	*P*	01	−02	−04	−03
17	*R*	100	04	−27	−01
18	*VbIQ*	37	40	00	−03
19	*L*	19	02	42	−11
20	*F*	05	−67	−25	53
21	*K*	28	38	39	04
22	*Hs*	21	−52	07	63
23	*D*	14	−49	03	64
24	*Hy*	20	−29	48	53
25	*Pd*	−10	−08	03	86
26	*Pa*	−10	−33	−07	70
27	*Pt*	08	−69	−14	67
28	*Sc*	−01	−58	−22	74
29	*Ma*	−08	−05	−59	39
30	*Es*	28	77	09	−45
31	*A*	−02	−40	−09	83
32	*R*	−16	11	51	16

---

[2] Since the sample used was primarily composed of cases selected for a previous analysis of Rorschach determinants [10], the original correlations between Rorschach variables were utilized in the present matrix.

[3] The loadings on Factor V were so low that it was not included in later matrices.

Table 4

Oblique Factors

Variable No.	Symbol	I″	II″	III″	IV″
1	*W*	33	82	−12	−23
2	*D*	91	36	−48	04
3	*d*	65	41	−35	−10
4	*M*	60	42	−36	−06
5	*FM*	48	21	−39	−01
6	*m*	37	30	−49	11
7	*k*	43	60	08	−15
8	*K*	10	59	07	01
9	*F*	86	33	−37	−05
10	*Fc*	21	41	−12	−03
11	*c*	51	55	−17	−13
12	C^1	18	32	−26	−06
13	*FC*	50	36	−48	06
14	*CF*	27	72	12	−18
15	*C*	02	46	06	−11
16	*P*	06	−06	−35	−03
17	*R*	99	60	−44	−02
18	*VbIQ*	27	54	08	−18
19	*L*	16	15	20	−11
20	*F*	11	−47	−12	76
21	*K*	16	51	33	−12
22	*Hs*	19	−19	25	77
23	*D*	09	−18	40	78
24	*Hy*	11	02	54	60
25	*Pd*	−18	−05	09	82
26	*Pa*	−11	−27	−04	77
27	*Pt*	13	−45	−09	89
28	*Sc*	01	−45	−19	90
29	*Ma*	−05	−11	−39	37
30	*Es*	19	71	−09	−73
31	*A*	−05	−24	17	91
32	*R*	−17	09	55	10

Table 5

Correlations between Oblique Factors

Factor	I″	II″	III″	IV″
I″	1.00			
II″	38	1.00		
III″	−42	15	1.00	
IV″	−05	−26	−18	1.00

a general Rorschach productivity factor. A similar factor was found in previous studies, but in the course of rotation much of its variance was spread among other factors. As can be seen, rotation did not appreciably alter the factor in the present study. This dimension accounts for a high percentage of the total variance in the intercorrelation matrix, and its presence indicates that other factors must be treated conservatively.

Factors II and III are bipolar factors indicating that relationships between the two tests as they are now scored are probably inverse. That is, in Factor II, many of the MMPI scales have high negative loadings, particularly *Pt* and *Sc,* while the factor has positive loadings on Rorschach *W, CF,* and the *K* variables, Verbal IQ, and MMPI *Es* and *K.* The oblique and orthogonal positions are quite divergent. The peak loadings on the two solutions are similar, but the oblique axis includes a wider range of Rorschach determinants and greater productivity. As shown in Table 5, the latter solution results in correlation between Factors I and II.

From the loadings on Factor II, it seems reasonable to postulate that patients who are productive, giving a wide range of Rorschach determinants, who can organize the blots into wholes, and who use color with some freedom will tend to deny concern with themselves, dissatisfaction, pessimism, and worry, thus achieving a high *Es* score on the MMPI. In this case, Factor II may be tentatively labeled "ego strength."

Factor III is also bipolar, including significant loadings on MMPI *R* (repression), *L, K, D,* and *Hy,* along with Rorschach *F.* Its opposite pole is made up of *Ma* and Rorschach *R, D, M, FM, m,* and *FC.* Perhaps it may best be labeled "expressive-repressive."

Factor IV, a general MMPI dimension, is closely related to the first factor obtained in the several factorial studies mentioned above. It seems to involve a set to say psychiatrically negative things about oneself.

Summary and Conclusions

1. MMPI, Rorschach, and Wechsler-Bellevue data on 100 psychiatric patients were selected at random from the files of the Clinical Psychology Section at Letterman Army Hospital.

2. The intercorrelation matrix was factor analyzed by the centroid method.

3. Four factors were obtained and described. While relationship between the two tests was not remarkable, there is evidence for an inverse one. Two factors demonstrated overlap between the two tests. Factor II, tentatively

called "ego strength," included high loadings on the MMPI "ego-strength scale," and a fairly wide range of Rorschach determinants, with negative loadings on the MMPI "complaint" scales. Factor III contained high loadings on the MMPI "repression" scale and Rorschach *F*, with negative loadings on the movement variables, *FC* and the MMPI *Ma* scale.

Factors I and IV were identified respectively with Rorschach productivity and the MMPI "general maladjustment" factor. They indicate that a considerable amount of variance in each test is not accounted for by covariation between them.

4. The data suggest that a Rorschach variable might be derived that would function as a reliable predictor of ego strength.

5. The MMPI findings are in agreement with results of previous factorial studies.

Received August 17, 1953.

References

1. Altus, W. D. Some correlates of the group Rorschach and the schizophrenic scale of the group MMPI among two groups of "normal" college students. *J. consult. Psychol.*, 1948, **12**, 375–378.
2. Barron, F. An MMPI scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327–333.
3. Cattell, R. B. *Factor analysis.* New York: Harper, 1952.
4. Clark, J. Some MMPI correlates of color responses in the group Rorschach. *J. consult. Psychol.*, 1948, **12**, 384–386.
5. Ellis, A. Recent research with personality inventories. *J. consult. Psychol.*, 1953, **17**, 45–49.
6. Thompson, G. M. MMPI correlates of certain movement responses in the group Rorschachs of two college samples. *J. consult. Psychol.*, 1948, **12**, 379–383.
7. Tyler, F. T. A factorial analysis of fifteen MMPI scales. *J. consult. Psychol.*, 1951, **15**, 451–456.
8. Welsh, G. S. A factor study of the MMPI using scales with item overlap eliminated. *Amer. Psychologist*, 1952, **7**, 341-342. (Abstract)
9. Wheeler, W. M., Little, K. B., & Lehner, G. F. J. The internal structure of the MMPI. *J. consult. Psychol.*, 1951, **15**, 134–142.
10. Williams, H. L., & Lawrence, J. F. Further investigation of Rorschach determinants subjected to factor analysis. *J. consult. Psychol.*, 1953, **17**, 261–264.
11. Wittenborn, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1950, **14**, 261–267.
12. Wittenborn, J. R. Level of mental health as a factor in the implications of Rorschach scores. *J. consult. Psychol.*, 1950, **14**, 469–472.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

The Diagnosis of Subclinical Schizophrenia[1]

Donald R. Peterson
University of Illinois

One way to refine a scheme of psychiatric classification is to examine patients who, when certain developments in the course of their illness have been considered, turn out to have been incorrectly diagnosed. Careful study of behavior recorded at the time of their diagnoses should permit more definite limitation of class boundaries and more accurate placement of other patients for whom a diagnosis is required.

This study is concerned with the behavior of a group of veterans who should have been diagnosed as "latent," "incipient," or "subclinical" schizophrenics, but who in fact received some other designation. The importance of making the classification correctly for the protection of other members of society, for promoting maximal economy in treatment, and generally for facilitating wise patient disposition should be obvious.

Subjects and Procedures

Group 1. A set of false negatives was obtained by searching the complete file of case records at a VA mental hygiene clinic for all male, white patients who met the following criteria:

a. They were examined by a psychologist and were seen for at least two interviews by staff members at the clinic.

b. They received some diagnosis which contained no derivative of the word "schizophrenia."

c. They were later sent to a psychiatric hospital and there were said to be schizophrenic.

Group 2. A sample of true negatives was selected by examining data previously recorded [4] for a sample of nonhospitalized patients at the same clinic. Each false negative in Group 1 was considered in turn, and was compared consecutively with each patient listed in the data for the nonhospitalized sample until one was found who had received a primary diagnosis identical with that given his counterpart in Group 1. For these cases, neither diagnosis nor outpatient status was subsequently changed. The sample was again limited to male, white veterans.

Group 3. An exhaustive sample of true positives was found by searching the clinic files and selecting those patients who, as in the other groups, had received psychological and psychiatric examination at the clinic, whose diagnoses were "latent," "incipient," or "subclinical" schizophrenia, or "schizophrenia in remission," and who later were sent to the hospital and received some variety of the diagnosis "schizophrenia" there. Once more, all cases were male and all were white.

Groups 1 and 2 each contained 33 cases. For Group 3, $N = 27$. Data were gathered with regard to age, education, Wechsler-Bellevue IQ, and Minnesota Multiphasic Personality Inventory (MMPI) profile, and the false negatives compared with each of the two control samples by means, chiefly, of a χ^2 analysis, as specified by McNemar [2, formula 87] for 2 by k tables. This method was employed for testing differences in education and for investigating the discriminatory effectiveness of six signs said by Meehl [3] to be commonly employed in differentiating MMPI psychotic profiles from neurotic ones. The signs, modified only to make objective counting possible, are as follows:

1. T scores on four or more of the clinical scales over 70
2. $F > 65$
3. $Sc > Pt$

---

[1] Published with the permission of the Chief Medical Director, Department of Medicine and Surgery, Veterans Administration, who assumes no responsibility for the opinions expressed or the conclusions drawn by the author.

4. *Pa* or $Ma > 70$
5. *Pa* or *Sc* or $Ma > Hs$ and *D* and *Hy*
6. $D > Hs$ and *Hy*

In addition, the *t* test was applied for all continuous variables [1, formula 5.03]. Where an *F* ratio suggested dissimilarity in population variance, a method for approximating *t* was used which involves no assumption of homogeneous variance [1, pp. 74-75].

Results and Discussion

The diagnoses actually received by the false negative group are presented in Table 1.[2] They seem to have been most often confused with cases of anxiety reaction or mixed psychoneurosis.

Table 1
Primary Diagnoses of Patients in the False Negative Group

Diagnosis	Number
Psychoneurosis, mixed or unclassified	9
Anxiety reaction	9
Conversion reaction, hysteria or somatization reaction	4
Obsessive-compulsive reaction	3
Inadequate personality	1
Nonpsychiatric diagnoses	7

Differences in age, education, and intelligence were not significant either in the comparison between Groups 1 and 2 or that between Groups 1 and 3. Examination of *T* score means for the various MMPI scales suggested that the false negatives differ from the members of Group 2 on six of the scales, while they differ from the members of Group 3 with regard to three scales. Results of this type are not revealing enough to warrant interpretation.

Pattern analysis, even of the simplest kind, seems to yield far more meaningful information. A striking aspect of the profiles presented in Fig. 1 is the essential similarity between test patterns of the latent schizophrenics who were not correctly diagnosed and those who were. When Meehl's psychotic signs were tested, not one produced a significant χ^2. In terms of the test pattern, however, the patients in Group 1 are obviously very much *unlike* the sample of true negatives, even though, as a group, they had been given identical diagnoses. Every one of the six signs yielded a χ^2 significant at or beyond the .05 level. In view of the informal way in which the criteria were established, the results are surprisingly good.

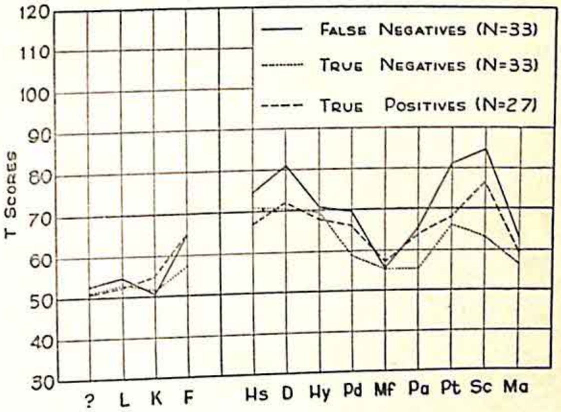


Fig. 1. Mean MMPI profiles.

After-the-fact examination of the profiles leads me cautiously to suggest utility for one sign not included in this list or any other I know. It is the relationship between *Hy* and *Pd*—in inferential statement, a relationship between the repressive, inhibitory tendencies of the patient, and his propensity for "acting out" whatever impulses he may have. The "significance" which it yields here has little meaning, but the sign does not appear to be mentioned in other studies [4, 5] and might be worth including in future investigations of this type.

The implications furnished by the data are clear. Adequate information was there for the diagnosticians to use; it just was not employed as effectively as it might have been. By assigning to all cases one point for the presence of each sign, it is possible to derive a total "psychotic" score for any patient. When this was done, and a cutting score for class designation placed between one and two, it was found that 88 per cent of the incorrectly diagnosed patients in Group 1 would have been accurately placed. This increase, however, would have been gained at the expense of incorrect diagnosis for 39 per cent of the cases in Group 2. A cutting score between two and three would have led to correct designation for 67 per cent of the members of Group 1, and incorrect diagnosis for 18 per cent of the Group 2 cases. The over-all percentage of accurate classification would have been 74 for both cutting points.

---

[2] A table of means, variance estimates, and results of significance tests for age, IQ, and the MMPI scales has been deposited with the American Documentation Institute. Order Document No. 4203 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.25 for 35 mm. microfilm or $1.25 for 6 x 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

These figures are far short of ideal, but they offer some hope for nosologic differentiation that is considerably better than that which has been made before. Groups 1 and 3 are exhaustive samples of their respective patient classes, and consideration of their relative sizes suggests that the diagnosis of subclinical schizophrenic is missed more often than it is correctly made.

That the problem with which we deal is one of inaccurate diagnosis, and not one of essential change in the patients themselves, is suggested by the results of follow-up scrutiny of 19 members of Group 1 who were retested at the hospital. Neither configural analysis of the type described above nor a trend analysis of the direction of sign movement produced statistically reliable results. Either the patients did not change much or the MMPI is too insensitive an instrument to measure any changes that did occur.

Summary

Refinement in psychiatric classification can be effected through specification of the behavioral characteristics of groups of false positives and false negatives. By isolating cases who are found, in the light of later developments, to have been incorrectly diagnosed, and by measurement of the behavior they exhibited before these facts were known, it is possible to improve diagnostic precision.

The present study is concerned with the diagnosis of subclinical schizophrenia, an important but difficult diagnosis for the clinician to make. Thirty-three patients had been examined at an outpatient clinic, and had received there some diagnosis containing no form of the word "schizophrenia." They were later hospitalized and were designated schizophrenic by staff members at the institution.

Comparisons were made between this group and another group of clinic patients, matched for initial diagnosis, with regard to age, education, Wechsler-Bellevue IQ, and performance on the MMPI. Similar comparisons were made between the criterion group and an exhaustive sample of 27 outpatients who, like the first group, were later hospitalized and diagnosed schizophrenic, but for whom the diagnosis had been correctly made at the clinic.

No differences in age, education, or intelligence were found. Neither did any clearly interpretable results emerge from analysis of the MMPI by means of the *t* test. Configural analysis through application of a group of psychotic signs, however, indicated that the MMPI patterns of the false negatives were essentially like those of the patients for whom the diagnosis, subclinical schizophrenia, had been accurately applied, and were substantially different from those produced by the group of patients with whom they had originally been classed. Diagnosis made in terms of these signs would have led to a considerable increase in classificatory precision.

Received August 14, 1953.

References

1. Johnson, P. O. *Statistical methods in research.* New York: Prentice-Hall, 1949.
2. McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.
3. Meehl, P. E. Profile analysis of the MMPI in differential diagnosis. *J. appl. Psychol.,* 1946, 30, 517–524.
4. Peterson, D. R. Predicting hospitalization of psychiatric outpatients. *J. abnorm. soc. Psychol.,* 1954, 49, 260-265.
5. Wiener, D. N. A control factor in social adjustment. *J. abnorm. soc. Psychol.,* 1951, 46, 3–8.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

The Relationship between Anxiety Level and Several Measures of Intelligence

Joseph D. Matarazzo, George A. Ulett, Samuel B. Guze, and George Saslow
Division of Psychosomatic Medicine and Departments of Medicine and Neuropsychiatry, Washington University School of Medicine

In an earlier study we reported the results of an experiment involving learning as a function of increasing levels of anxiety [4]. The study used 101 college sophomore *S*s, who were divided into seven groups, from low to high anxiety level, according to each individual's score on the Taylor Scale of Manifest Anxiety [8]. Individuals with Taylor raw scores between 1 and 5 were given an anxiety rating of 1, while those with raw scores of 6-10, 11-15, 16-20, 21-25, 26-30, 31-35 were given ratings of 2, 3, 4, 5, 6, and 7, respectively. There was one *S* with a score of 40, who was included with those given an anxiety rating of 7. Analysis of the Taylor raw score distribution yielded, reading from 1 to 7, the following distribution of *S*s along the seven-point scale: 4, 23, 27, 15, 17, 7, and 8. The results indicated that human maze learning had a significant curvilinear relationship to increasing levels of anxiety when total time to learn the maze was the criterion, and a significant rectilinear relationship when number of trials was the criterion of learning. Thus, when time was the measure of learning, *S*s in the middle anxiety range were superior to *S*s in both the high and low anxiety groups, whereas with the trials measure, *S*s at the high end were inferior to *S*s in the middle and low ends.

In an effort to evaluate whether differences in anxiety level were in fact responsible for the differences found in the mean learning abilities of the seven groups or whether the obtained differences were due to differences in intelligence levels of the groups, two available criteria of intelligence were used. Each *S* had been given the 1949 edition of the American Council on Education (ACE) Psychological Examination for College Freshmen, furnishing one criterion. The other criterion was each *S*'s grade-point average for the previous year, his freshman year.

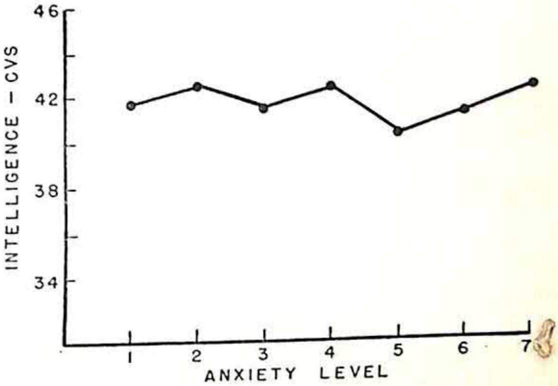


Fig. 1. Mean grade-point average for each of seven groups ranging from low anxiety (group 1) to high anxiety (group 7).

When the grade-point criterion was plotted as a function of anxiety level, no differences were found; i.e., grades had no significant relationship to level of anxiety. This can be seen in Fig. 1 where the mean grade-point averages for each of the seven groups are plotted. The function is essentially a straight line parallel to the abscissa. The Pearson r for these data, an insignificant value of —.08, confirmed the graphic impression. The value of the unbiased correlation ratio, epsilon, [5, pp. 319-324] was .00, and the chi-square test of linearity of regression [5, p. 319] based on this value and the corresponding Pearson r was .00, indicating that the function does not, in fact, deviate from a straight line. Analysis of variance yielded an insignificant F value of .88. The use of t tests based on the pooled variance revealed only one significant difference, groups 4 and 5 at the .05 level, and since 21 tests were performed this can be considered a chance occurrence. Thus the conclusion of no relationship between grades and anxiety level appears warranted.

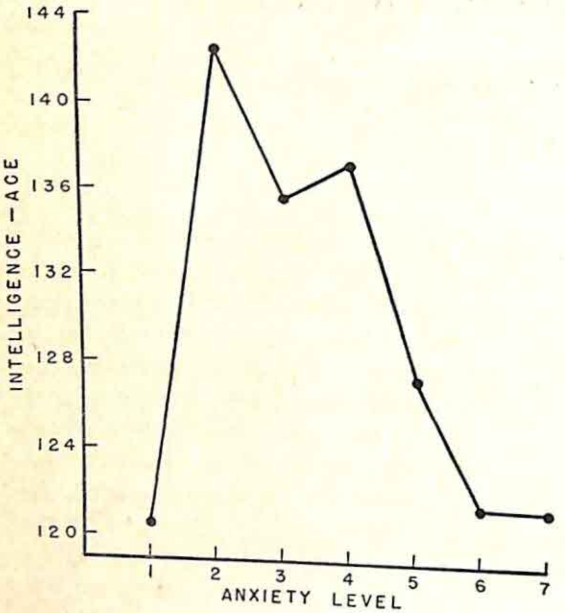


Fig. 2. Mean ACE score for each of seven groups ranging from low anxiety (group 1) to high anxiety (group 7).

The second criterion of intelligence, ACE scores, was analyzed next, as shown in Fig. 2. The apparent curvilinear relationship between total ACE score and increasing level of anxiety is strikingly similar to the actual function obtained for the *time* measure of learning and anxiety level in the earlier study [4]. However, unlike the time measure data of the earlier study, extensive statistical analysis revealed that the relationship between ACE score and anxiety level was in fact a rectilinear and not curvilinear one. Thus, the Pearson r between ACE score and anxiety level was —.25, a value significant at the .01 level of confidence, indicating a slight but nevertheless significant negative relationship between ACE score and anxiety level. The value of epsilon of .25 (.01 level of confidence) is equal to the value of r, and therefore the value of the chi-square test of linearity is .00, indicating a rectilinear rather than a curvilinear relationship.

As was pointed out in our earlier study, analysis of variance is limited as a test of differences among means with data of this type, for the reasons pointed out by Webb and Lemmon [9, 10]. Thus, with the ACE data of Fig. 2, we obtained an F value of 1.90, which is short of the 2.17 value needed for significance at the .05 level. Despite this lack of significance for F, two-tailed t tests based on the pooled variance yielded four significant differences: groups 1 and 2 at the .05 level and groups 2 and 5, 2 and 6, and 2 and 7 each at the .02 level of confidence. These values indicate that we are dealing with true differences rather than chance occurrences. However, the limitations of analysis of variance when one variable is ordered (our anxiety groups) should be kept in mind. Correlational techniques, like those reported above, are thus the methods of choice.

As with the data of our earlier study, the question of the small N (four Ss) in group 1 was next raised. It is obvious that the N's unduly influence the nature of the curve shown in Fig. 2, since the chi-square test results indicate that the relationship is a rectilinear one, i.e., that the essentially straight line from group 2 to 7 is probably a better indication of the relationship between ACE score and anxiety level. In an effort to test this, the data were regrouped as in the earlier study in an effort to increase the N in group 1. Since the 101 Ss had a range of Taylor raw scores from 1 to 40, it was decided that grouping them into five new groups in intervals of eight Taylor points (1–8, 9–16, 17–23, 24–32, 33–40) would be justifiable. The regrouping now placed 13 Ss in group 1, 47 in group 2, 23 in group 3, 14 in group 4, and 4 in group 5. Thus, while group 1 changed from 4 Ss to 13, group 5 was reduced to 4. Since it was felt that the N in group 1, and not those at the high end of the anxiety scale, was crucial, the 4 Ss in group 5 were included with the 14 Ss in group 4, thus yielding four new groups.

Figure 3 presents the relationship between ACE scores and these four differing levels of anxiety. As was predicted from the relationship between the r and epsilon obtained for the seven groups, the use of broader categories better represents the relationship between anxiety level and ACE score. Statistical analysis bears this out.

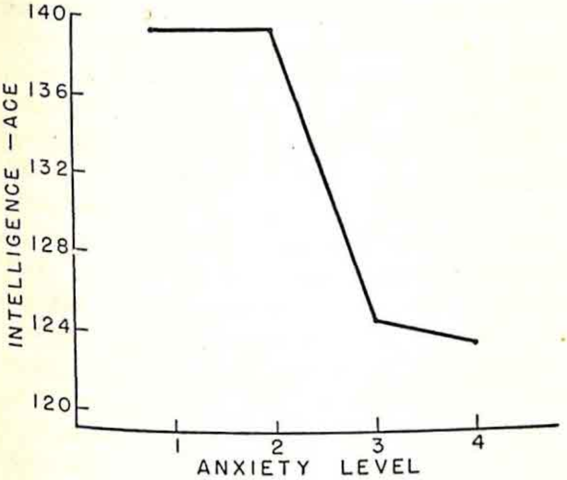


Fig. 3. Mean ACE score for each of four groups ranging from low anxiety (group 1) to high anxiety (group 4).

The value of r for the data of Fig. 3 was, of course, still —.25 (significant at the .01 level). The new value of epsilon was .28 (also significant at the .01 level). As the slight difference between these correlations would indicate, the value of the chi-square test of linearity of regression, 2.91, did not approach significance. Analysis of variance yielded an F of 4.03 (.01 level), and analysis by t tests indicated that groups 1 and 2 were *each* significantly different from groups 3 and 4: groups 1 and 3 and 1 and 4 at the .05 level; groups 2 and 3 and 2 and 4 at the .01 level of significance. These analyses and those of Fig. 2 indicate that the relationship between ACE score and anxiety level is rectilinear, and the curve in Fig. 3 is the best representation of this relationship. Therefore it can be concluded that anxiety appears to be moderately, although significantly, correlated with total ACE score and that the higher the anxiety level the lower, on the average, is the obtained score on this test. Analysis of the two parts of the ACE revealed that the relationship between the Q and L scores and anxiety was approximately the same. Both yielded negative correlations comparable to the one shown in Fig. 3 for the total score.

Since there was a lack of agreement between the relationship of anxiety level and our two original criteria of intelligence, an explanation was sought. The Pearson r between total ACE score and grades was .51, significant at the .01 level. As this value is very near the mode of all reported correlations between these two measures [7, p. 120], it was not believed that our sample was responsible for the lack of agreement between the relationship obtained for anxiety and each of our two criteria. Instead the answer was sought in the difference between our two criteria.

As the ACE is a timed test, we wondered if the speed factor could account for the negative correlation obtained with increasing levels of anxiety. This hypothesis was strengthened by two further considerations: (a) our earlier study revealed a significant relationship between anxiety level and the time measure of learning, and (b) clinical experience indicates that anxious *S*s do not function efficiently under time pressure. We therefore decided to investigate this hypothesis by retesting *S*s, this time with a nontimed intelligence test. If the obtained function is a straight line parallel to the abscissa, i.e., shows no relationship between this measure of intelligence and anxiety level, we can conclude that the hypothesis that the time factor of the ACE accounts for its negative correlation with anxiety level is reasonable.

The nontimed intelligence test used was the *CVS*, an abbreviated version of the Wechsler-Bellevue developed by Hunt and his associates [3]. None of the three subtests of this scale, Comprehension, Vocabulary, and Similarities, is timed, and *S*s were allowed, in a group administration, to take as much time as they required to complete the scale.

The results, when this criterion of intelligence (total *CVS* score) is plotted against increasing levels of anxiety, are given in Fig. 4. The data are shown as a function of all seven groups. As the figure demonstrates, and actual analysis revealed, the use of the broader categories (four groups) gave an essentially similar curve. Inspection of Fig. 4 indicates that the hypothesis that the time factor of the ACE may be responsible for the obtained negative correlation appears to be borne out. Statistical analysis confirms this impression, as the values

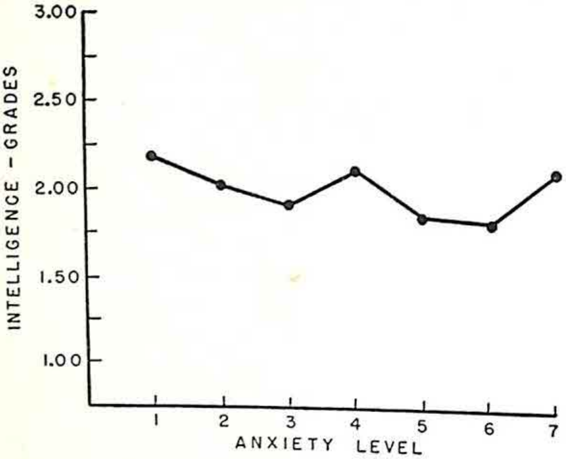


Fig. 4. Mean *CVS* score for each of seven groups ranging from low anxiety (group 1) to high anxiety (group 7).

of *r*, epsilon, and chi square[1] were —.07, —.14, and —1.72. None of these values approached significance. The value of *F*, .68, was insignificant, and *t*-test analyses revealed only one, and therefore chance, significant (.05 level) difference out of the 21 tests computed. Analysis of the relationship of *C*, *V*, and *S*, taken separately, as a function of anxiety level revealed three curves similar to Fig. 4.

It can therefore be concluded that intelligence, as measured by the *CVS*, like grades as a measure of intelligence, is not related to anxiety level. However, the relationship between *CVS* score and grades is not to be construed as a noncomplex affair, as the relatively low, although significant (.01 level) Pearson *r* of .34 between them demonstrates. This value is typical of those reported between measures of intelligence and grades [7, p. 145]. Finally, the Pearson *r* between ACE and *CVS* scores of .57 (.01 level) is, again, the typical finding [7, pp. 117-118]. The lack of even stronger relationships among these various measures indicates the complexities introduced by the interaction of personality and intellectual variables, as well as the limitations of validity and reliability inherent in the criteria themselves.

The finding in this study of a slight but nevertheless significant negative correlation between total ACE score and increasing levels of anxiety is consistent with the results of several recent studies. In a series of publications, Sarason and his associates have studied the relationship between anxiety, measured by a questionnaire sampling reactions to examinations and intelligence tests, and various learning, achievement, and psychometric tasks. In one of these studies Sarason and Mandler [6] demonstrated that low-anxious *S*s performed significantly better on two college entrance examinations (Mathematical Aptitude Test and Scholastic Aptitude Test) than did *S*s scoring at the high-anxious end. As in our study, a high level of anxiety was associated with lower performance on these two aptitude tests. Furthermore, these authors found no significant relationship between their anxiety measure and actual grade average for the previous year's work — a finding consistent with ours. And finally, their obtained significant correlation of .21 (.05 level) between anxiety level and a 20-minute timed intelligence test, the Henmon-Nelson Test of Mental Ability, college level, is again consistent with our correlation of .25 (.01 level) between anxiety level and the timed ACE scale. Thus, these two studies, completed independently of each other and using different indices of anxiety as well as different measures of intellectual functioning, yielded essentially similar results.

Two recent Rorschach studies report results consistent with these findings. In a controlled experiment, Beier showed that the effect of induced anxiety, a standardized Rorschach personality interpretation designed to promote stress, was to significantly decrease performance on such intellectual tasks as abstract reasoning, sorting, and mirror tracing [1]. It is of interest that several of Beier's measures were timed. In the other experiment, Gaier obtained an index of anxiety-readiness for each of his 11 *S*s on the basis of their Rorschach protocols and related this characteristic to each *S*'s performance on a variety of scholastic and other type tests [2]. He found a rank-order correlation between anxiety-readiness and the *L* score of the ACE of —.37, and —.39 between the *Q* score and anxiety-readiness. These results are again similar to those of our study.

---

[1] The negative sign results from the chance negative value of epsilon. This is discussed by Peters and Van Voorhis [5, p. 355].

The implication of these findings, and those of the present study, is not to be drawn from the magnitude of the correlations between a personality variable ("anxiety") and various measures of "intelligence." These correlations are too low. Rather, the importance of the findings lies in the fact that they are consistent with clinical observation, and they point out an area of fruitful research once new and more valid measuring instruments are devised. The interaction of personality and intellectual variables, so-called, will receive much needed study once the problem of objective measures has been solved, especially for personality. The Taylor scale is one small step in that direction.

Summary

Three criteria of intelligence were each evaluated as a function of increasing levels of anxiety for 101 *S*s who were divided into groups according to their scores on the Taylor Scale of Manifest Anxiety. The results indicated no relationship between anxiety level and the first and third of these criteria: grade-point average and score on the *CVS* scale. The relationship between the second criterion, score on the ACE scale, and increasing levels of anxiety was found to be a slight but nevertheless statistically significant (.01 level) negative one (r of —.25). The fact that the ACE is a timed scale appeared to be a reasonable hypothesis to account for the negative correlation obtained. The use of the third criterion, the untimed *CVS* scale, seemed to bear out this hypothesis.

Received September 14, 1953.

References

1. Beier, E. G. The effect of induced anxiety on flexibility of intellectual functioning. *Psychol. Monogr.*, 1951, **65**, No. 9 (Whole No. 326).
2. Gaier, E. L. Selected personality variables and the learning process. *Psychol. Monogr.*, 1952, **66**, No. 17 (Whole No. 349).
3. Hunt, W. A., French, Elizabeth G., Klebanoff, S. G., Mensh, I. N., & Williams, M. Further standardization of the CVS individual intelligence scale. *J. consult. Psychol.*, 1948, **12**, 355–359.
4. Matarazzo, J. D., Ulett, G. A., & Saslow, G. Human maze performance as a function of increasing levels of anxiety. *J. gen. Psychol.*, in press.
5. Peters, C. C., & Van Voorhis, W. R. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill, 1940.
6. Sarason, S. B., & Mandler, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 810–817.
7. Super, D. E. *Appraising vocational fitness.* New York: Harper, 1949.
8. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
9. Webb, W. B., & Lemmon, V. W. A qualification in the use of analysis of variance. *Psychol. Bull.*, 1950, **47**, 130–136.
10. Webb, W. B., & Lemmon, V. W. A sequel to the notes of Patterson and Diamond. *Psychol. Bull.*, 1952, **49**, 155.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

Comparison of Two Forms of the Manifest Anxiety Scale

Joyce B. McCreary and A. W. Bendig[1]
University of Pittsburgh

Taylor [1] reported the development of a Manifest Anxiety Scale (MAS) consisting of 50 items drawn from the MMPI. Scores are obtained either by scoring only these items from the full-length MMPI or by embedding them in 175 additional filler items and giving the resulting 225-item test. Taylor [1, p. 287] mentions that distributions of scores obtained by these two methods are not similar, and the correlation between scores from the two forms was only .68 ($N = 282$). Commonly, the 225-item test is used to identify anxious subjects, but this long test is not economical of either the subject's or the experimenter's time. Preferably only the 50 scored items should be given to *S*s. However, the results previously quoted suggest that scores from this shortened form might not be comparable to those from the longer forms.

Students ($N = 56$) were given two forms of the MAS. Form A contained the 225 items described above, but only 50 of them, scattered irregularly throughout the form, were scored. Form B contained only the 50 scored items from Form A, all other items being omitted. In one section ($N = 22$), Form A was given first followed immediately by Form B. In the other section ($N = 34$), the forms were administered in reverse order.

The 112 scores were converted to normalized T scores ($M = 50$, $SD = 10$), based upon the performance of the norm group of 1,971 students at the University of Iowa [1]. Analysis of variance showed that the variation between classes (which was confounded with the order of presentation of the two test forms) and between forms was not significant. Differences between the trials were significant at the .05 level of confidence. Comparison of trial means indicated that *S*s tended to show lower (less anxious) scores on the second test given them regardless of form. The correlation between Form A and Form B transformed scores was .95: a similar correlation between pairs of raw (untransformed) scores was .96.

The shortened form of the MAS (Form B) gave results almost identical with those obtained from the long form (Form A). Evidently either form can be used for the selection of anxious *S*s with equivalent results. Since Form B is more parsimonious in terms of class time, its use is preferable. The significant between-trials term in our analysis of variance suggests an explanation of the differences found by Taylor between scores from the full-length MMPI and those from the 225-item MAS. Many of the MAS items are near the end of the MMPI, and *S* has had to answer several hundred other items before reaching them. The "fatigue" effect found in our study may be accentuated and compounded when an even longer test is used.

Brief Report
Received February 8, 1954.

Reference

1. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285–290.

---

[1] An extended report of this study may be obtained without charge from A. W. Bendig, Department of Psychology, University of Pittsburgh, Pittsburgh 13, Pa., or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4211 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.25 for 35 mm. microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

Rater Experience and the Reliability of Case History Ratings of Adjustment

A. W. Bendig and Janine Sprague
University of Pittsburgh

In Guilford's discussion of the optimum length of rating scales [5, pp. 267-268], the hypothesis is advanced that longer scales, with seven to nine categories, should be used only with experienced and well-motivated judges, whereas shorter scales should be used with relatively naive judges. Presumably, longer scales place too great a burden on the discriminatory ability of inexperienced raters. Although widely adopted as a guiding principle in the construction of rating scales, this rational hypothesis has, as far as we have been able to determine, never been subjected to experimental test. Previous research [2, 3] has confirmed several of Guilford's strictures on constructing rating scales and has questioned others. The present research was designed to clarify the relationships between scale reliability, the past experience of the rater, and length of the scale. Rater experience is used here as task-relevant experience, i.e., the amount of past exposure to and familiarity with the kind of judging task required of the rater.

Procedure

Stimuli. Ten short (approximately 100-word) case histories were abstracted from several commonly used textbooks in abnormal psychology. The cases ranged from mildly disturbed neurotic to institutionalized psychotic. Most of the cases were reduced in length, with the more technical wording either eliminated or simplified. The cases were mimeographed as a booklet with a face sheet of instructions to the raters. The instructions requested the rater (*a*) not to discuss the cases with his friends before rating, (*b*) to make his ratings as objective as possible, and (*c*) to make judgments independent of previously rated case histories.

Scales. Numerical rating scales with 3, 5, 7, or 9 categories were mimeographed on separate sheets with spaces for recording the ratings of the case histories. Three simple verbal anchors were used, one under each of the two end scale categories and one under the center. The anchors were:

Slightly maladjusted
Moderately maladjusted
Extremely maladjusted

Subjects. The total group of 208 raters was composed of four experience groups: 108 students in introductory psychology; 62 in the second semester of abnormal psychology; 29 graduate students in psychology; and 9 psychologists having the Ph.D. degree and several years experience in clinical or counseling work. Most of the introductory students were freshmen or sophomores, and they rated the cases sometime after the period in the course devoted to personality and behavior disorders. Almost all of the abnormal psychology students were juniors and seniors, with a large minority majoring in psychology. The graduate students ranged from first-year to fourth-year students, with a strong majority in clinical and counseling areas. All of the Ph.D. raters were affiliated with this university or some neighboring agency in a clinical or counseling capacity and had a minimum of three years' experience in service activities. For the first three groups, the case histories and rating sheets were passed out in class, and students were permitted to rate outside of the classroom. The Ph.D. raters were contacted individually and also rated at their leisure. Scales with 3, 5, 7, and 9 categories were randomly distributed to the introductory and abnormal psychology students. Because of the smaller numbers of raters, and 5- and 7-category scales were randomly distributed to the graduate students, and only the 7-category scale was given to the Ph.D.'s.

Table 1

Reliability Analysis of Adjustment Ratings

Rater group	*N*	Number of categories	Rater reliability			Rater bias	
			F	Group *R*	Individual *r*	*F*	*R*
Introductory psychology students	27	3	15.64**	.94	.35	.76	—.31
	26	5	17.76**	.94	.39	1.15	.13
	21	7	19.68**	.95	.47	2.17**	.54
	34	9	39.97**	.97	.53	2.69**	.63
Abnormal psychology students	13	3	15.20**	.93	.52	3.44**	.71
	15	5	12.99**	.92	.44	2.76**	.64
	18	7	29.83**	.97	.62	4.67**	.79
	16	9	16.60**	.94	.49	4.79**	.79
Graduate students	13	5	18.76**	.95	.58	2.25*	.56
	16	7	23.94**	.96	.59	2.32**	.57
Clinical Ph.D.	9	7	17.65**	.94	.65	2.07*	.52

* Significant at the .05 level.
** Significant at the .01 level.

Results

The reliability of the individual raters for each of the 11 subgroups was estimated by analysis of variance techniques [2, 3]. These results can be found in Table 1.

In general, individual rater reliability increased with higher levels of education, with the Ph.D. group showing the largest reliability ($r = .65$). Consideration of the interaction of rater experience and length of scale discloses an interesting phenomenon. Introductory students show a smooth linear increase in reliability as the scale lengthens until, with the 9-category scale, they surpass the abnormal psychology students using a 9-category scale and almost reach the reliability level of the graduate students. The four subgroups of abnormal psychology raters fluctuate in reliability, showing their highest reliability with the 7-category scale and their lowest with 5 categories. This greater fluctuation may be due, in part, to the smaller numbers of *S*s per group. However, the general reliability trend is upward for longer scales, but the rate of increase (slope of the best-fitting straight line) is not as steep for the abnormal as it is for the introductory psychology *S*s. There was no difference in the rater reliability found for the two graduate student subgroups using 5- and 7-category scales, and both groups showed slightly smaller reliability than did the Ph.D. group.

As to the magnitude of constant errors in the ratings of the subgroups (rater bias), the introductory psychology *S*s show a progressive increase with longer scales, surpassing the Ph.D. group in bias with a 7-category scale, but not surpassing the graduate student subgroups until the 9-category scale was reached. The abnormal psychology raters consistently showed a greater bias for every scale length, but again showed some fluctuation between these subgroups. Of all four subgroups using the 7-category scale, the Ph.D. raters showed the smallest bias.

The mean ratings of the ten case histories by the 11 subgroups were intercorrelated and the matrix of coefficients found.[1] All of the 55 correlations were extremely high (the median being .90), with 54 significant at the .01 level, $df = 8$, and the remaining correlation, $r = .73$, significant at the .05 level. A centroid factor analysis was performed on these correlations, replicating the analysis twice to stabilize the communalities; the unrotated factor loadings can be found in Table 2. Factor extraction was terminated after two factors because of small residual correlations. Factor I was a broad general factor accounting for an average of 91 per cent of the variation in mean

---

[1] A one-page table showing intercorrelations between mean adjustment ratings of eleven groups of raters has been deposited with the American Documentation Institute. Order Document No. 4206 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.25 for 35 mm. microfilm or $1.25 for 6 x 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Table 2

Factor Loadings from Analysis of Intercorrelations between Mean Adjustment Rating of Rater Groups

Group	Number of categories	Factor* I	II	h^2
Introductory psychology students	3	94	—37	102
	5	95	—24	96
	7	97	—18	97
	9	95	—23	96
Abnormal psychology students	3	90	15	83
	5	95	17	93
	7	96	20	96
	9	97	15	96
Graduate students	5	95	26	97
	7	97	22	99
Clinical Ph.D.	7	97	18	97
Per cent of total rater variance		91	5	96

* Decimal points omitted.

case history ratings and appeared to be a function of the length of the scale. Subgroups using a 7-category scale consistently showed higher loadings on Factor I, and the two subgroups using 3-category scales had the smallest loadings. Factor II was associated with 5 per cent of the remaining intragroup rating variance and was clearly related to the experience background of the raters. This "experience" factor was bipolar, with the four introductory subgroups having negative loadings and the other subgroups spaced along the positive side of the vector. In all, these two factors accounted for 96 per cent of the variance between mean case history ratings.

The nonparametric rank test of Kruskal and Wallis [6] was used to test the above interpretation of I as a scale-length factor and II as an experience factor. The 11 subgroups were ranked according to their loadings on each of the factors, and the sums of ranks of the four introductory, four abnormal, and three graduate student-Ph.D. subgroups on Factor I and on Factor II were found. Applying the rank test to the sums of ranks of the three experience groups on Factor I gave an *H* value of 1.30, which is not significant at the .05 level, while applying the same test to the three sums of ranks on Factor II yielded an *H* of 8.91, which is significant at the .01 level [6, p. 615]. Again the 11 subgroups were ranked on each factor and the sums of ranks for each of the four lengths of scale obtained. The *H* test for Factor I was 7.48 (significant at the .05 level) and for Factor II, *H* equalled 3.17 (not significant at the .05 level). These tests of significance support our interpretation of Factor I as related to scale length and Factor II as associated with rater experience. Objection may be raised to our interpretation of Factor II since our experience groups did show a general increase in rater reliability from introductory students to Ph.D. clinicians. The interpretation of Factor II as a "reliability" factor does not appear plausible when we note the large overlap in rater reliability between the introductory and abnormal student groups (Table 1) and the complete lack of overlap between these groups on Factor II.

Finally, the above Factor II results suggested the hypothesis that the introductory-abnormal-graduate student groups should show a progressive increase in the correlation of their mean case history ratings with those of the Ph.D. group. The average correlations of the two introductory, two abnormal, and two graduate student groups who used the 5- and 7-category scales with the Ph.D. group were found and tested for significant differences by the *r*-to-*z* technique described by Edwards [4, pp. 132-136]. The critical ratios resulting from these comparisons can be found in Table 3. The introductory group showed a significantly smaller average correlation with the Ph.D. group than did the abnormal or graduate student groups, on a one-tailed test corresponding to the above directional hypothesis. However, the comparison of average correlations of the abnormal and graduate student groups with the Ph.D. group showed no significant difference.

Table 3

Significance of the Differences between Average Correlations of Three Rater Groups with the Clinical Ph.D. Raters

Group	*N*	$\bar{r}$	Critical ratios Abnormal	Graduate
Introductory	47	.87	1.88*	3.37**
Abnormal	33	.96		1.07
Graduate	29	.97		

* Significant at the .05 point.
** Significant at the .01 point.

Discussion

The results of this study appear sufficiently unequivocal to question seriously the Guilford-sponsored statement that longer scales should be used only with experienced raters. Our results suggest just the opposite: for more reliable ratings longer scales should be used with inexperienced raters and (tentatively) short scales can only be resorted to when the raters have had experience in the type of judgment required. As a rationale to explain our results we suggest that inexperienced raters are more likely to disagree as to the scale placement of a particular stimulus than are experienced raters and disagree to the extent, approximately, of one scale category. With long scales this one-category disagreement represents a much smaller percentage of the total rating variance than with short scales. Since rater reliability is a function of the ratio of rater disagreement variance to the total variance, we would expect shorter scales, with smaller total variance, to yield less reliable ratings. However, experienced raters can better tolerate the compression of the scale and better maintain agreement with short scales.

Parenthetically it might be noted that our results are a tribute to the rating ability of the Ph.D. clinicians. Since we wanted to compare ratings from all levels of the experience continuum, both our case histories and scale anchors had to be designed to be comprehensible to the lowest common denominator — the introductory psychology student. Consequently, many of the Ph.D. clinicians justifiably felt that our case histories were too short and incomplete for them to make adequate judgments and that the points on our scales were inadequately defined. In spite of these limitations the average intercorrelation among the nine Ph.D. clinicians was .65, and their group judgment had a reliability of .94. Doubtless with more complete case histories and with a better definition of what we meant by "adjustment," these reliabilities would have been increased.

The factor analysis of the intercorrelations of mean ratings indicates that the largest influence in these correlations is the length of the scale. Factor I, which is fairly clearly related to scale length, accounts for 91 per cent of the total variability in the mean ratings. Factor II was even more clearly an experience factor, but only tapped 5 per cent of the rating variance. We must admit at this point that the difference between our major experience groups, which is related to Factor II, is composed of a number of different variables. Presumably there are differences in age, intelligence, educational training, motivation, and sex percentages between introductory, abnormal, graduate student, and Ph.D. groups. Which of these organismic variables are most responsible for differences in rating can only be a matter of speculation at this point. However, the academic selection processes that operate to winnow from introductory psychology a small group who will receive the Ph.D. also operate to increase the agreement among the raters as to what constitutes maladjusted behavior. This selective process appears to be most influential between the introductory and abnormal psychology courses and is probably related to whatever factors determine that students will become majors in psychology.

The increasing reliability of naive raters with longer scales found in this study contrasts markedly with the results of previous studies [2, 3]. All three of these studies required self-rating on the part of the rater, i.e., using the scale to report the rater's perceptions of his own feelings, preferences, etc. In the former studies, scale reliability did not markedly increase with longer scales. In the present research, the raters were asked to evaluate stimuli that were more objective and more independent of the rater, i.e., verbal descriptions of the behavior of other people. Presumably these ratings could still be affected by a projective mechanism, but it seems likely that somewhat less ego-involving stimuli would encourage greater interjudge agreement [1]. We would predict that naive or inexperienced judges would show a steeper gradient of increasing rater reliability with lengthened scales if even more objective stimuli were rated.

Summary

Raters ($N = 208$) judged the adjustment level shown in each of ten short case histories. Four rater groups were used: introductory psychology, abnormal psychology, psychology graduate students, and a group of post-Ph.D.

clinical psychologists. Rating scales with 3, 5, 7, or 9 categories were used. Results indicated a general increase in reliability with higher educational levels. Rater reliability increased sharply with longer scales for the introductory group and increased at a slower rate for the abnormal psychology group. Factor analysis of the intercorrelations among the groups indicated that a factor related to the length of the scale accounted for 91 per cent of group variance, and a second factor reflecting experience differences measured 5 per cent of the variance.

Received August 17, 1953.

References

1. Bendig, A. W. Inter-judge vs. intra-judge reliability in the order-of-merit method. *Amer. J. Psychol.*, 1952, **65**, 84–88.
2. Bendig, A. W. The reliability of self-ratings as a function of the amount of verbal anchoring and of the number of categories on the scale. *J. appl. Psychol.*, 1953, **37**, 38–41.
3. Bendig, A. W. Reliability and the number of rating scale categories. *J. appl. Psychol.*, 1954, **38**, 38–40.
4. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
5. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
6. Kruskal, W. H., & Wallis, W. A. Use of ranks in one-criterion variance analysis. *J. Amer. statist. Ass.*, 1952, **47**, 583–621.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

Personality Correlates of Undergraduate Field of Specialization[1]

Richard C. Teevan
University of Michigan

This paper attempts to investigate whether or not the personality of an individual is correlated significantly with his choice of major field in college. In recent years there have been two research programs which stimulated the present work: the survey of scientists by Knapp and Goodrich [4] and the investigations of eminent scientists by Roe [5, 6]. These studies attempted to explore the personalities and backgrounds of men who are presently working in the sciences.

Since the above investigations, among others, have shown that there is a correlation between personality and vocation, the question arises whether this correlation can be demonstrated before the vocation is actually entered or whether the vocation itself fits the personalities of its members into a common mold. The present paper throws light on this matter by using college students who have already expressed their preference for a major field but who have not been in that field long enough for it to affect their personalities. The fact that choice of a major in college is not necessarily followed by a similar vocation is a drawback, but one that might serve only to attenuate any obtained relationships.

Procedure

The diagnostic instrument in this study was the Blacky Pictures [2], a modified projective technique measuring psychosexual development. Eighty-five male students from Wesleyan University, ranging in age from 17 to 25 years, were tested. The subjects were chosen by their major field, with approximately the same number from each of the academic divisions. The major fields were broken down into three divisions: 1. music, art, German, French, and English ($N = 29$); 2. government, economics, and history ($N = 28$); and 3. geology, astronomy, biology, physics, and chemistry ($N = 28$). An attempt was made to obtain subjects who were sincerely interested in their field of endeavor, so that an average of B— was taken as the criterion.

The Blacky Pictures were administered in three sections, each section, insofar as possible, containing equal numbers of subjects from each division. There were two deviations from standard procedure: (*a*) Up to 15 seconds longer than the standard two minutes were allowed on the spontaneous stories; (*b*) the inquiry questions, rather than being flashed on the screen one at a time with a time limit for each question, were printed at the bottom of each page. Subjects were instructed not to glance at them until told to. An observational check and questioning at the end of the session showed good observance of these instructions.

Results and Discussion

The findings of this study are summarized in Tables 1 and 2. Table 1 gives the significance of the differences between divisions when Blacky disturbance scores are related to choice of major field. In Table 2 the percentages of each group showing disturbed and neutral responses are given for those Blacky dimensions on which significant differences were obtained.

The first division (English, art, languages) was found to be the highest of the three divisions on Oral Eroticism. Since the great majority of the subjects tested (22 out of 29) were from the English department, it is perhaps possible to treat them as literary persons. Bergler [1] observed 36 writers in psychoanalytic treatment and concluded that all writers

---

[1] This paper is abstracted from an honors thesis done at Wesleyan University under the guidance of Professor Robert H. Knapp.

Table 1

The Relationship between Blacky Scores of Disturbance and Choice of Major Field

Blacky category	*p*	Ranking of division by degree of disturbance		
		Most	Middle	Least
Oral Eroticism	.01	1	2	3
Oral Sadism	.05	2	1	3
Anal Expulsiveness	NS			
Anal Retentiveness	NS			
Oedipal Intensity	.06	2	1	3
Masturbation Guilt	NS			
Castration Anxiety	NS			
Identification Process	NS			
Sibling Rivalry	NS			
Guilt Feelings	.01	2	1	3
Ego Ideal	NS			
Narcissistic Love Object	NS			
Anaclitic Love Object	.07	2	1 = 3	1 = 3

Note.—Division 1—music, art, German, French, English; Division 2—government, economics, history; Division 3—geology, astronomy, biology, physics, chemistry.

Table 2

Percentage of Each Division Showing Disturbed (+) and Neutral (O) Responses to Blacky Dimensions

Blacky dimension	Division		
	1 Lit.	2 Soc.	3 Sci.
Oral Eroticism			
Disturbed (+)	61	31	18
Neutral (O)	39	69	82
Oedipal Intensity			
Disturbed (+)	57	66	36
Neutral (O)	43	34	64
Oral Sadism			
Disturbed (+)	46	55	18
Neutral (O)	54	45	62
Guilt Feelings			
Disturbed (+)	64	86	39
Neutral (O)	36	14	61
Anaclitic Love Object			
Disturbed (+)	11	31	11
Neutral (O)	89	69	89

are psychic masochists and therefore fixated at the oral erotic level. Boyd [3], using the Blacky Pictures on children, found more disturbance at the oral erotic level in good readers than in retarded readers. Thus the present findings from the Blacky Pictures are consistent with previous descriptions of literary people as passive and dependent.

The second division (economics, government, and history) was highest on the other four categories that showed differences: Oral Sadism, Oedipal Intensity, Guilt Feelings, and Anaclitic Love Object. Three of these categories contain a common element — emphasis on the mother. Thus Oral Sadism is related to aggression toward the mother (see Blacky Picture No. 2); Oedipal Intensity indicates an unresolved conflict over desire for the mother; and Anaclitic Love Object reflects a desire for the mother as a love object. Guilt Feelings, the fourth dimension on which the second division is highest, is based on the introjection of parental standards and also conceivably fits into this pattern. Thus the disturbances prevalent in Division 2 tend to center around the boy's relationship to his mother. Any interpretation of this set of findings is sufficiently speculative to allow the reader the freedom of his own projections in this regard. The facts themselves, however, seem quite clear.

Another striking point is the fact that Division 3 (the sciences) has the lowest disturbance scores on most of the Blacky categories. Since no artifacts can be uncovered to account for this finding, we can conclude that college students who select a major field within this division are less disturbed than those who major in the other fields used in this study.

Summary

This investigation was conducted to determine whether or not personality factors are correlated with choice of a major field in college. Three broad groupings of college majors were compared on scores derived from the Blacky Pictures [2]. The following significant differences were found:

1. The "literature" division had higher disturbance scores on Oral Eroticism than the other divisions.

2. The social sciences division had higher disturbance scores on Oral Sadism, Oedipal Intensity, Guilt Feelings, and Anaclitic Love Object.

3. The science division had the lowest disturbance scores on nearly all categories.

From the above results it is concluded that correlations between personality and vocation previously found by Roe [5, 6] for professional groups can also be demonstrated during the period preceding entrance into a profession.

Received November 5, 1953.

References

1. Bergler, E. *The writer and psychoanalysis.* New York: Doubleday, 1950.
2. Blum, G. S. A study of the psychoanalytic theory of psychosexual development. *Genet. Psychol. Monogr.,* 1949, **39**, 3-99.
3. Boyd, R. D. Reading retardation as related to personality factors of children and their parents. Unpublished doctor's thesis, Univer. of Michigan, 1953.
4. Knapp, R. H., & Goodrich, H. *Origins of American scientists.* Chicago: Univer. of Chicago Press, 1952.
5. Roe, Anne. A Rorschach study of a group of scientists and technicians. *J. consult. Psychol.,* 1946, **10**, 317–327.
6. Roe, Anne. Psychological examinations of eminent biologists. *J. consult. Psychol.,* 1950, **13**, 225-246.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

Congruence of Self and Ideal Self in Relation to Personality Adjustment

Thomas E. Hanlon, Peter R. Hofstaetter, and James P. O'Connor
The Catholic University of America

Recent studies of patient change during successful psychotherapy [4, 6, 7, 8, 9] have demonstrated an increase in the correlation between the client's self concept and his ideal self concept as measured by Roger's *Q* sort.[1] The hypothesis of the present study springs directly from this observation. If successful psychotherapy promotes better adjustment, a positive and significantly high correlation should exist between measures of adjustment and the congruence of the self concept and the ideal self. This correlation should also exist in a sample from a population in which departures from satisfactory adjustment are still within the "normal" range, that is, in a population whose members are not patients. Studies by Bills *et al.* [1] and by Draffen [2] suggest that the hypothesis will be confirmed.

Procedure

The subjects of this study were 78 male juniors in a Catholic high school in Wilmington, Delaware. The mean age of the group was 16.3 years, with a range from 14 to 18 years and a standard deviation of .78 years. The mean IQ, as measured by the Otis Self-Administering Tests of Mental Ability, Intermediate, Form B, was 104.3 with a range from 77 to 140 and a standard deviation of 11.4.

Since the *S*s were available for only short periods of time, measures of the self concept and ideal self concept were obtained by employing a modification of the rather lengthy *Q*-sort technique. The 100 self-referent statements were presented on one sheet of paper, with five categories ranging over the qualitative scale from "most characteristic" to "least characteristic." After the experimenter (T.E.H.) gave appropriate verbal instructions that differentiated the two concepts of self, the *S*s sorted the statements into the five categories, first according to their self concept and then according to their ideal self concept. The number of statements to be placed in each category were 6, 25, 38, 25, and 6, respectively, thus forcing the items into a binomial distribution.

The last step in the testing procedure was the administration of the California Test of Personality, Secondary Series [10], appropriate for students from the ninth grade to the college level. Two measures of adjustment, personal and social, are given by this test. The total score may be regarded as an indicator of over-all adjustment.

The self concept and the ideal self concept were correlated (Pearsonian r) and the resulting coefficients were transformed into Fisher's z to facilitate statistical treatment. The individual z values were then correlated with the percentile scores obtained from the California test.

The first and the fourth quartiles, each comprising 20 *S*s, were compared with respect to specific areas and specific items of the California test.

Results and Discussion

The individual correlations between self concepts and ideal self concepts ranged from —.34 to .79; the corresponding z values were from —.35 to 1.07. The mean z was .42 and the standard deviation was .26. A comparison of the z distribution with the normal curve, by the area method, yielded a χ^2 of 4.89 for 6 df; the corresponding p value was higher than .50. Consequently, the distribution of z values can be considered normal. Retransformation into Pearson r's gave a sample mean of .40. The fiducial limits at the .01 level for the mean of the universe were .33 and .46, and the lower limit of .33 at the .001 level was still significantly greater than zero. It may be concluded that the correlations between the self and the ideal self were distributed normally around a positive mean value.[2]

---

[1] For the sake of brevity, the correlation between self concept and ideal self concept will be referred to as "self-ideal congruence." Both concepts of self are operationally defined by the results of respective *Q* sorts.

The correlations between self-ideal congruence (z) and the percentile scores of the California test and its subscales are given in Table 1, together with the correlation with IQ. The right-hand columns of Table 1 show a comparison between two extreme groups with regard to self-ideal congruence, i.e., Ss in the first and fourth quartiles of the z distribution. With one exception, the correlations between self-ideal congruence and the adjustment scores are significant at least at the .05 level. The differences between the means of the two extreme groups are also significant. It is not surprising that the correlation involving "social standards" (subscore 2A) is not significant, for it is indeed doubtful whether mere knowledge of social standards reflects adjustment in general or, more specifically, social adjustment.

Table 1

Correlations between Self-Ideal Congruence, Adjustment Scores, and Intelligence

Variable	Correlation with self-ideal congruence (z)	Mean score: High self-ideal congruence	Mean score: Low self-ideal congruence	t
Total adjustment	.70*	84.50	34.25	9.29*
1. Self adjustment	.70*	80.50	36.25	6.57*
A. Self reliance	.48*	73.70	47.50	4.09*
B. Sense of personal worth	.50*	80.00	43.30	4.95*
C. Sense of freedom	.40*	69.00	36.75	4.78*
D. Feeling of belonging	.55*	75.50	40.25	5.65*
E. Freedom from withdrawal tendencies	.56*	77.00	29.40	6.00*
F. Freedom from nervous symptoms	.39*	74.50	46.65	3.42**
2. Social adjustment	.59*	80.25	33.00	7.13*
A. Social standards	.08	70.25	60.05	1.20
B. Social skills	.52*	70.90	30.00	5.59*
C. Freedom from antisocial tendencies	.22***	75.45	51.40	2.71**
D. Family relations	.30**	85.75	57.60	3.24**
E. School relations	.45*	65.90	31.50	4.53*
F. Community relations	.50*	76.75	27.20	5.33*
IQ	.05	105.80	105.95	0.05

* Significant at the .001 level.
** Significant at the .01 level.
*** Significant at the .05 level.

The F test of linearity [11, p. 278] was applied to the correlation between self-ideal congruence and total adjustment (.70), and the resulting value of .86 was found to fall far short of that required for the significance of deviations from rectilinearity at the .05 level, i.e., an F of 2.14 for 7 *df* and 69 *df*. Thus, the correlation between self-ideal congruence and total adjustment is not only marked but also rectilinear. It is concluded, therefore, that the congruence between the self concept and the ideal self concept can be used as a measure of adjustment with a considerable degree of confidence.[3] This confirmation of the basic hypothesis of the present study also lends justification to the manner in which the Rogerian school has employed correlations between self-sorts in the investigation of the psychotherapeutic process.

The correlations between IQ and self-ideal congruence (.05), IQ and total adjustment (.12), IQ and self adjustment (.11), and IQ and social adjustment (.11) were all found to be not significantly different from zero. These findings indicate that self-ideal congruence and adjustment are not a function of intelligence. This result is in substantial agreement with the findings of Kimber [3], who has reported a correlation of .04 between the Army Alpha scores and California test scores of 400 college students.

Although there was a slight tendency for the older *S*s to score lower on measures of adjustment and self-ideal congruence, *t* tests showed that no significant relationship was present.

---

[2] This statement is also true of the distribution of Pearson r's, which is negatively skewed, but not significantly deviant from normality ($\chi^2 = 8.34$ for 5 *df*; $p > .10$). However, for purposes of averaging and subsequent correlational analysis, Fisher's z's are preferable.

[3] The regression equation between the r values of self-ideal congruence and the percentile scores for total adjustment (TA) reads: $TA = 26 + 87\,(r) \pm 17.3$. The possible values of r which satisfy this equation thus range from —.50 to 1.00.

A comparison of the *S*s in the first and fourth quartiles of the *z* distribution (self-ideal congruence) was made on the basis of their responses to the 180 items of the California test. The two groups showed no difference in their manner of responding to 12 of the items. For the remaining items, the low self-ideal congruence group have a higher percentage of unfavorable responses to 155, while the high self-ideal congruence group gave relatively more unfavorable responses to only 13. In the latter instance, none of the differences was statistically significant. On the other hand, of the 155 items which showed a greater percentage of unfavorable responses for the low self-ideal congruence group, 42 of the differences were significant at the .05 level and 14 at the .01 level.

A qualitative analysis of the items that discriminated between the two groups showed that *S*s in the low self-ideal congruence group were characterized by a lack of persistence in work habits and by feelings of oppression and inadequacy. They were more likely to feel rejected by both family and friends and tended to react to this by withdrawing within themselves. For this reason, their fantasy life seemed to be more active than that of the high self-ideal congruence group. The social standards and tendency to avoid antisocial activities were the same for both groups, but *S*s of the low self-ideal congruence group showed much more uneasiness in social situations. They tended to stay in the background and to avoid contacts calling for poise and diplomacy, for they were easily hurt and were aware of their proneness to quarrel. They tended to dislike school, and their negativistic attitudes made for difficulties in their relationships with teachers, fellow students, and members of the opposite sex.

It is clear from the above description that the personality features of psychotherapy patients are likely to coincide with the characteristics of *S*s showing a low congruence between self concept and ideal self concept.

Summary

The relationship between measures of adjustment and the congruence of the self concept and ideal self concept in a sample of 78 high school students was investigated, using the California Test of Personality and a modified *Q*-sort technique. The following conclusions were drawn:

1. The correlation between the self concept and the ideal self concept tends to be positive, with a mean value in an unselected population between the fiducial limits of .33 and .46. Therefore, the over-all tendency is toward a congruence of the two concepts of self.
2. Congruence between self and ideal self is a normally distributed trait.
3. The correlation between self-ideal congruence and total adjustment is positive and highly significant with regression being rectilinear. Therefore, the use of measures of self-ideal congruence in evaluating the extent of personality maladjustment appears justified.
4. The hypothesis which underlies the use of *Q* sorts in evaluating change in psychotherapy is confirmed.
5. Intelligence and age show no significant relationship with self-ideal congruence and with measures of adjustment.
6. Maladjustment in a person need not require that his self concept be negatively related to his ideal self. Where the correlation is minimal (r less than .27), signs of maladjustment may already be manifest.

Received September 8, 1953.

References

1. Bills, R. E., Vance, E. L., & McLean, O. S. An index of adjustment and values. *J. consult. Psychol.*, 1951, **15**, 257-261.
2. Draffen, Mary T. The relation between self-self ideal consistency and inconsistency and personality traits. Unpublished master's thesis, The Catholic Univer. of America, 1951.
3. Kimber, J. A. The insight of college students into the items on a personality test. *Educ. psychol. Measmt*, 1947, **7**, 411–420.
4. Mowrer, O. H. *Q*-technique—description, history and critique. In O. H. Mowrer (Ed.), *Psychotherapy: theory and research.* New York: Ronald, 1953. Pp. 316-375.

5. Raimy, V. C. Self reference in counseling interviews. *J. consult. Psychol.*, 1948, **12**, 153–163.
6. Rogers, C. R. Perceptual reorganization in client-centered therapy. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: an approach to personality*. New York: Ronald, 1951. Pp. 307-327.
7. Rogers, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.
8. Rogers, C. R. Studies in client-centered psychotherapy III: the case of Mrs. Oak — a research analysis. *Psychol. Serv. Center J.*, 1951, **3**, (1-2), 47-165.
9. Seeman, J., & Raskin, N. J. Research perspectives in client-centered therapy. In O. H. Mowrer (Ed.), *Psychotherapy: theory and research*. New York: Ronald, 1953. Pp. 205-234.
10. Tiegs, E. W., Clark, W. W., & Thorpe, L. P. The California test of personality. *J. educ. Res.*, 1941, **35**, 102–108.
11. Walker, H. M., & Lev, J. *Statistical inference*. New York: Holt, 1953.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

Validity of a Critical-Item Scale for Schizophrenia on the MMPI[1]

Harold Rubin
VA Hospital, Lebanon, Pa.

In an attempt to eliminate the problem of "false positives" which confronts many clinicians using the MMPI, Benarick [1] has developed a critical-item scale to aid in differentiating psychotics from nonpsychotics. The scale, reported earlier in this journal [2], consists of 11 items taken from the *Sc* scale of the MMPI and scored in the *Sc*-plus direction. Benarick also presents a second scale consisting of 10 items answered more frequently in the *Sc* direction by nonpsychotics. The present study attempts to test the validity of these critical-item scales for schizophrenia on a hospital population. It differs somewhat from Benarick's original study in that the nonpsychotic group consists of hospitalized subjects.

Procedure

MMPI profiles selected were those of 64 male patients of a veterans hospital who were administered the MMPI as part of an admission test battery, and whose raw *Sc* score was greater than 20. Thirty-two of the patients were diagnosed as schizophrenics with a mean raw *Sc* score of 29.6, and thirty-two were diagnosed as psychoneurotics with a mean raw *Sc* score of 28.3. All diagnoses were made by the psychiatric staff of the hospital following a period of observation.

Results and Discussion

As may be observed in Table 1, the mean scores on the eleven-item scale were 3.8 for the psychotics and 2.9 for the nonpsychotics. The highest number of critical items obtained by any single subject was eight, this being true in both groups.

Table 1

Distributions of Scores of Psychotics and Nonpsychotics on Two Critical-Item Scales

Score	Eleven-Item Scale		Ten-Item Scale	
	Psychotic	Nonpsychotic	Psychotic	Nonpsychotic
10				
9			1	2
8	1	1	1	1
7	2	0	5	7
6	2	1	7	4
5	7	5	6	8
4	7	3	6	5
3	5	8	2	2
2	3	6	3	2
1	5	6	1	1
0	0	2	0	0
Total	32	32	32	32
Mean	3.8	2.9	5.0	5.3

Following Benarick's [2] suggestion, a cutting score of 4 was used to identify psychotics from nonpsychotics. With this criterion, it was found that 59% of the psychotics were correctly identified, while 31% of the nonpsychotics yielded "false positives." For the total group of 64 patients, 62% of the identifications made were correct. In an earlier study, Rubin [4] has shown that psychotics tend to have their highest scores on the *Sc* scale. On this basis, if one were to assume that each profile represented that of a psychotic, 50% of this group of 64 would be correctly identified. Benarick's scale offers a 12% improvement in such correct identifications.

A chi square computed between psychotics and nonpsychotics, using the cutting score of 4, yielded a value of 5.11, which is significant at less than the 5% level. Although not as significant as the findings reported by Benarick [2], some validity, it seems, may be attributed to this critical-item scale. However, when chi-square values were computed separately for each of the eleven items, not a single item was found to differentiate significantly between psychotics and nonpsychotics. The distribution of scores among the eleven items was markedly similar for both groups. One exception was that on three of the items (76, 157, and 331) the nonpsychotics obtained a slightly greater, although not significant, frequency. This is in contrast to the findings of Benarick, who reported significant differences between psychotics and nonpsychotics on the individual items as well as on the total scale.

---

[1] From the Veterans Administration Hospital, Lebanon, Pennsylvania.

On the ten-item scale, as observed from Table 1, it was found that the distributions for both psychotics and nonpsychotics were so similar that further statistics were unnecessary. It should be stated, however, that among both psychotics and nonpsychotics there was a tendency for those who obtained a high score on the eleven-item scale to obtain a high score on the ten-item scale. This would indicate that the scales are not as complementary as one might expect from Benarick's results.

The findings of the present study would appear to indicate that even though it was possible to combine a number of items into a scale having some diagnostic value when used on a group level, the items taken singly failed to be statistically meaningful in separating psychotics from nonpsychotics. Since only four of the critical items need be scored to make a differential diagnosis or to detect "false positives," which, after all, is the purpose of the scale, it would seem that the value of the scale in differentiating between psychotics and nonpsychotics is limited, particularly when the subjects are hospitalized patients. The present study also indicates the difficulty of constructing a meaningful scale based on psychiatric diagnoses which are themselves of questionable reliability [3].

Received September 30, 1953.

References

1. Benarick, S. J. An investigation of the schizophrenic scale of the Minnesota Multiphasic Personality Inventory. Unpublished master's thesis. The Pennsylvania State College, 1950.
2. Benarick, S. J., Guthrie, G. M., & Snyder, W. U. An interpretive aid for the *Sc* scale of the MMPI. *J. consult. Psychol.*, 1951, **15**, 142-144.
3. Rotter, J. B. Clinical methods: psychodiagnostics. *Annu. Rev. Psychol.*, 1953, **4**, 295-316.
4. Rubin, H. The MMPI as a diagnostic aid in a veterans hospital. *J. consult. Psychol.*, 1948, **12**, 251-254.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

Wechsler-Bellevue Performance of Delinquent and Nondelinquent Girls

Julia R. Vane
Hempstead Public Schools
and Virginia W. Eisen
Catholic Charities Guidance Institute

The purpose of this study was to determine whether a group of adolescent female delinquents would show a characteristic test pattern on the Wechsler-Bellevue Scale. Other studies concerned with delinquent patterning have dealt with male offenders, with one exception, a recent study by Bernstein and Corsini [4]. These authors investigated the effect of order of presentation of the subtests of the Wechsler-Bellevue Scale on the scores of female delinquents. Their hypothesis that the psychopathic pattern described by Wechsler is a function of the conditions under which delinquents are tested was not substantiated. They found that the rank order of the subtests was virtually the same for the two groups despite the difference in order of presentation.

The subjects in the present study included an experimental group of 100 delinquent females between the ages of 16 and 25, with IQ's between 70 and 110 on the Wechsler-Bellevue Scale, and a control group of 38 nondelinquent females in the same age and IQ range. All *S*s in both groups were white and native born. All were interviewed and given the Wechsler-Bellevue Scale by one of the authors.

The experimental group consisted of delinquents committed by the New York State courts to the House of Good Shepherd, Peekskill, New York, a reformatory for girls and women adjudged guilty of offenses such as prostitution, forgery, and shoplifting. Psychiatric and psychological interviews and case history material were used to eliminate delinquents suffering from psychoses, brain damage, alcoholism, or psychoneuroses. The great majority of the delinquents (93%) fell within the age range in which the greatest incidence of crime is reported, i.e., 16 to 21 years [8], and in this respect should be representative of delinquents in general.

The control group consisted of girls from two New York City public high schools. To avoid inclusion of possible delinquents, no girl was selected who was considered a behavior, personality, or social adjustment problem. Truancy, even of the shortest duration, was also a basis for exclusion, since it had been found that 99% of the delinquent group had records of truancy. School guidance cards, attendance records, and interviews with the school guidance counselor, as well as interviews with the girls themselves, were used to evaluate the variables.

The nondelinquents were equated as closely as possible with the delinquents for level of intelligence and chronological age. Years of schooling could not be matched since in order to obtain nondelinquents similar in age to the delinquents, it was necessary to choose the nondelinquents from the junior and senior years of high school. As the result all the nondelinquents had completed at least three years of high school, whereas 91% of the delinquents had left school during the first and second years of high school, if not before. Practically all of the nondelinquents were enrolled in vocational or nonacademic courses.

With respect to socioeconomic background, the delinquents and nondelinquents closely resembled one another. The majority of the girls from both groups resided in the same areas of New York City, and their fathers' occupations were similar. On the Goodenough-Anderson Occupational Rating Scale [6] 82% of the fathers' occupations in both the delinquent and the nondelinquent group fell in the lower three categories, namely, V—semiskilled, minor clerical, minor business, VI—slightly skilled, and VII—day laborers.

The majority of delinquents (53%) were the children of parents both born in the United States. Of the remainder, not more than 10% had parents who were born in any one particular country. Half of the nondelinquents (50%) also had both parents born in the United States. Of the remainder, 26% were the offspring of parents one or both of whom were born in Italy. No other country was the birthplace of more than 8% of the parents of the nondelinquents.

With respect to intelligence, the mean total IQ of the control group showed a slight superiority, but the difference was not statistically significant ($t = 1.14$).

The mean chronological age of the delinquents (18.5) was 1.2 years greater than the mean age of the nondelinquents (17.3). Despite the statistical significance of this difference ($t = 6.00$), it should not influence the results since Wechsler grouped 17-, 18-, and 19-year-olds into one unit in standardizing the test.

Procedure

Mean verbal, performance, and full IQ's, and mean weighted subtest scores were computed for each group, and scatter was determined in terms of these. Because of the narrow age range and the fact that the groups were equated for level of intelligence, it appeared unnecessary to convert weighted scores into z scores as suggested by Barnett [3] or into subtest IQ scores as suggested by Alimena [1]. Rather extensive experimentation was done with both methods and neither was found useful or of greater value than the use of weighted subtest scores.

Results and Discussion

Table 1 shows the results for the two groups. The largest subtest difference was the superiority of the nondelinquents over the delinquents in Digit Symbol ($t = 3.88$). Wechsler [9] has noted that unstable individuals do poorly on this test and attributes this to their difficulty in concentrating and applying themselves for any length of time to a task requiring persistent effort. It is the authors' experience that delinquent girls do show inability for sustained effort, especially in simple, relatively monotonous tasks such as the Digit Symbol test. They often verbalize resistance to such tasks. The delinquents in Glueck and Glueck's investigation also showed a significantly poorer score on Digit Symbol than the nondelinquents. Glueck and Glueck feel that the Digit Symbol test "involves concentration and persistency of effort, and is therefore a test in which impulsive, emotionally labile persons would not perform as well as those who are stable" [5, p. 207].

Table 1

Wechsler-Bellevue Scores of Two Groups

Item	Delinquents ($N = 100$)		Nondelinquents ($N = 38$)		
	Mean	*SD*	Mean	*SD*	t
CA	18.5	1.46	17.3	.89	6.00
Weighted score	9.12	1.10	9.37	1.02	1.25
Full IQ	95.2	8.40	96.8	6.92	1.14
Verbal IQ	93.6	9.72	93.1	7.68	.31
Performance IQ	98.5	10.08	100.9	8.92	1.35
Information	8.43	2.22	9.39	1.97	2.46
Comprehension	9.31	2.17	9.47	2.19	.31
Digits	8.24	2.42	7.53	1.67	2.03
Arithmetic	6.48	2.38	6.61	2.43	.28
Similarities	8.15	2.22	8.08	2.26	.16
P. Arrangement	10.16	2.42	10.18	2.70	.00
P. Completion	9.64	2.43	9.26	2.17	.89
Block Design	9.70	2.83	9.97	2.13	.16
Object Assembly	10.86	2.43	11.24	1.87	1.00
Digit Symbol	10.42	2.55	11.74	1.43	3.88

The second largest difference between the groups was on the Information subtest ($t = 2.46$). On this, the nondelinquents were also superior. Although this may be related to the fact that our delinquent girls had fewer years of schooling than the nondelinquents, it seems to be of diagnostic significance. It has been noted repeatedly by educators, psychologists, and court workers that delinquents rarely work to capacity in school and consequently experience failure and frustration. They often seem to try to relieve their feelings of frustration by truancy. The comparatively low Information score of the delinquents seems to be diagnostic in that it is a reflection of that aspect of the delinquent personality which operates to prevent him from taking advantage of his educational opportunities.

On the Digit Span subtest the delinquents were superior to the nondelinquents ($t = 2.03$). The difference between the two groups is actually due to the fact that the nondelinquents did poorly on this subtest, and not that the delinquents did well. This is indicated by the fact that the mean score of the nondelinquents on this test (7.53) is considerably below their mean score on the ten subtests (9.37) and is also considerably below Wechsler's standardization subtest mean (9.3) for this age group. It is difficult to explain this result. It is possible that the testing situation was more anxiety provoking to the nondelinquents than to the delinquents and that this anxiety served to lower the Digit Span scores of the former, but actually it was not noted that the nondelinquents appeared any more anxious than the delinquents. It is possible that this should not be considered a valid indicator until further verification is obtained.

On the Comprehension test there was very little quantitative difference between the delinquents and the nondelinquents, but there were qualitative differences. The delinquents frequently responded to the questions by statements such as "Do you want me to tell you what I *should* do or what I *would* do?" They occasionally spoiled a good answer to the question "What is the thing to do if you find an envelope in the street . . . " by adding a remark such as "I would probably open it to see if there was any money in it." Asked "Why are laws necessary?" several countered with "You tell me!" or to the question why a marriage license is required they replied "I wish I knew." Several indicated that they would notify an usher or manager if they were the first person to discover a fire while sitting in the movies, and then added "I would really probably run." In these replies we see that the delinquents are aware of the socially accepted standards, although they may fail to apply them to their own behavior.

In both the delinquent and nondelinquent groups, the Performance IQ was considerably higher than the Verbal IQ, the difference being statistically significant (delinquents, $t = 3.5$, and nondelinquents, $t = 4.1$). As previous investigators [4, 5, 7, 9] had considered superiority of Performance scores over Verbal scores a delinquent or psychopathic sign, the fact that our nondelinquent group showed this superiority was somewhat unexpected.

Several hypotheses were considered as possible explanations for this result. One was that because the control group was more heavily weighted with *S*s whose parents were born in Italy (26%) than is the normal population, this factor might have influenced the results. It has been found that on certain other tests "Italian children tend to rate significantly higher on the performance scale (Arthur Point) than on the Binet" [2, p. 6]. To check this hypothesis the scores of the nondelinquents of Italian background were compared with the scores of the nondelinquent group as a whole. It was found that the mean Verbal IQ of the girls with one or both parents born in Italy was 91.9, and the mean Performance IQ was 99.7. The difference between the two scores was 7.8, which is exactly the difference found between the mean Verbal and Performance IQ's of the total nondelinquent group. In addition, the percentage of girls with Italian-born parents whose Performance score exceeded their Verbal score by five or more points was compared with the percentage of girls of non-Italian born parents whose Performance score was superior by five or more points. No significant difference was found. It was therefore felt that the fact that the control group contained a large percentage of girls whose parents were born in Italy did not account for the result.

A second hypothesis was that the mean IQ's of both groups in the present study, four to five points lower than Wechsler's standardization mean IQ of 100, might have had some bearing on the difference between Verbal and Performance scores, i.e., that lower IQ groups may normally tend to do better on Performance than on the Verbal scale. To check this explanation, both the delinquent and nondelinquent groups were divided into two subgroups on the basis of high and low IQ. The superiority of Performance score over Verbal score remained statistically significant in each case, although the difference was somewhat smaller in the higher IQ subgroups.

A further possibility is that since our two groups belonged to a lower socioeconomic level than the population in general, this may have influenced the results. As we had no nondelinquent group of higher economic level, and as figures have not been reported from any comparable group, this hypothesis cannot be tested at the present time.

The most plausible conclusion seems to be that although Performance superiority frequently has been associated with delinquency, actually this may be a characteristic pattern of normals at this age level. The fact that Wechsler's standardizing data [9, p. 222] for age levels 16 to 19 years also show a significant superiority (approximately $t = 3.0$) of Performance scores over Verbal scores would appear to substantiate this conclusion.

It should be noted in addition that two other signs associated with delinquency are present in our nondelinquent group. The sign "Picture Arrangement plus Object Assembly greater than Block Design plus Picture Completion" is found in 63% of the delinquent group and 66% of the nondelinquent group, and Arithmetic is the lowest ranking of the subtests in each group. The fact that the three above-mentioned signs have been associated with delinquents [4, 5, 7, 9] may have arisen because the delinquents were not compared with a sufficiently similar group of nondelinquents. It has been demonstrated that these signs are found in delinquent test profiles, but our results show that they are not unique for delinquents and occur also in certain normal groups.

In general it would seem that the successful use of test signs in differential diagnosis varies with the circumstances. For instance, it may be possible to differentiate delinquents from other clinical groups, such as neurotics, psychotics, and organics, on the basis of their Wechsler-Bellevue profiles, but these signs would not be adequate to differentiate between the "normal" nondelinquent girl and the true delinquent. It would appear that the act of delinquency itself would be a sufficient means of differentiation. Our experience has shown, however, that this is not always a good criterion, for the reformatory has girls who cannot be considered truly delinquent. Some have been committed by punitive families for a single sexual indiscretion. Others who ran away from intolerable home conditions were committed "to keep them out of trouble"; girls of this type were not included in our delinquent group. In contrast to these, the records of demonstrated psychopaths often showed previous suspended sentences on more serious charges. Actually, our records indicate that truancy is a highly crucial factor in the picture of delinquency and in the history of the individual delinquent. Serious consideration of this factor and more adequate and psychologically oriented methods of handling the early truant would thus appear important.

While our results do show some quantitative and qualitative test differences between the delinquent and the nondelinquent, they still seem to fall short of the desired accuracy. It would appear that the Wechsler-Bellevue Scale is not an adequate means of differentiating delinquents from nondelinquents when age, intelligence, and socioeconomic background are similar.

Summary

A group of delinquent and a control group of nondelinquent girls, equated for age and intelligence and similar in socioeconomic background, were given the Wechsler-Bellevue Intelligence Scale to determine if the delinquents would give a characteristic pattern on this test.

There was a statistically significant difference between these two groups on only two of the subtests, viz., Information and Digit Symbol, the delinquents being inferior to the nondelinquents on both. On the whole, the delinquents and the nondelinquents were more alike than different, showing similar deviations from Wechsler's norms for their age group. Three of the signs offered by Wechsler and Levi [7, 9] as characteristic of male psychopaths were found in both the delinquents and the nondelinquents in the present study.

On the basis of present findings it is concluded that female delinquents do show certain deviations from the standardization norms in respect to subtest variation on the Wechsler-Bellevue Scale. These deviations cannot, however, be considered unique to the delinquent since they also appear in nondelinquents of similar age, socioeconomic, and educational level. One might hypothesize that age, socioeconomic, educational, and cultural factors play a larger part in determining qualities of test performance on the Wechsler-Bellevue Scale than do deviant personality factors, but this hypothesis has yet to be adequately demonstrated.

Received October 16, 1953.

References

1. Alimena, B. Norms for scatter analysis on the Wechsler intelligence scales. *J. clin Psychol.*, 1951, **7**, 289–290.
2. Arthur, Grace. *A point scale of performance tests.* Vol. 1. *Clinical manual.* (2nd Ed.) New York: The Commonwealth Fund, 1943.
3. Barnett, I. The use of z-scores in equating the Wechsler-Bellevue subtests. *J. clin. Psychol.*, 1950, **6**, 184–188.
4. Bernstein, Rachel, & Corsini, R. J. Wechsler-Bellevue patterns of female delinquents. *J. clin. Psychol.*, 1953, **9**, 176–179.
5. Glueck, S., & Glueck, Eleanor. *Unraveling juvenile delinquency.* New York: The Commonwealth Fund, 1950.

6. Goodenough, Florence, & Anderson, J. E. *Experimental child study.* New York: Century, 1931.
7. Levi, J. A psychometric pattern of the adolescent psychopathic personality. Unpublished doctor's dissertation, New York Univer., 1943.
8. Tappan, P. W. *Juvenile delinquency.* New York: McGraw-Hill, 1949.
9. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 18, No. 3, 1954

Note on Sanders and Pacht's Investigation of Size Constancy of Clinical Groups

Edward Lovinger [1]
Michigan State College

In an investigation of perceptual size constancy of clinical groups, Sanders and Pacht [2] employed a rather unusual size-constancy index. The purpose of this note is to point out the differences between their index and those more commonly utilized and to demonstrate how its use is likely to be misleading.

They reported mean constancy indices of 105 for a group of normals, 99 for a group of neurotics, and 88 for a group of psychotics. The technique they employed for measuring constancy was one in which the comparison stimuli were far and the standard near.[2] The comparison stimuli consisted of 13 black squares graduated in size and randomly arranged under a consecutive series of numbers. These squares were always present at a distance of five meters from the subject. The standard squares were of five different sizes and were presented individually at two distances from the subject, first at four meters and then at two meters. The subjects were instructed to choose the comparison square that was equal in size to the standard square.

The following equation was utilized in obtaining the size-constancy index:

$$\text{Size-constancy index} = \frac{\text{Size}_{\text{near}} - \text{Size}_{\text{far}}}{\text{Size}_{\text{near}}}.$$

$\text{Size}_{\text{near}}$ and Size_{far} refer to the judged size of the standard at two and four meters respectively. In order to eliminate negative scores and decimals, a constant of one was added to the obtained ratio, and this result was multiplied by 100. Thus, perfect constancy would be represented by an index of 100. An index greater than 100 would indicate that the comparison square selected for the far standard (four meters) was smaller than the comparison square selected for the near standard (two meters), i.e., that the judged size of the standard was smaller when it was far than when it was near. An index of less than 100 would indicate that the judged size of the far standard was larger than that of the near.

Although this method of computation results in an index which clearly indicates the comparative judged sizes of the standard at the two distances, this index is likely to lead to misinterpretations if it is accepted as equivalent to a Brunswik or Thouless ratio. Sanders and Pacht at one point in their discussion seem to have done this. The authors stated: "The present results suggest findings which are similar to those of Thouless . . . , namely, that psychotics show less size constancy than normals" [2, p. 443]. This statement suggests not only a seeming misinterpretation of their own data, but also a misunderstanding of Thouless' experiment. Thouless [3] investigated the differences in constancy between cyclothymes and schizothymes. He found no significant differences between the two groups for three different constancy measures, although all the insignificant differences were in the direction of less constancy (underconstancy) for the schizothymes. Since there was no indication that any of these subjects were psychotic, including the schizothymes, it would seem rather unwarranted to speak of Thouless' results as indicating less size constancy for *psychotics* than for normals.

More important, however, is the implication that the authors' constancy index is equivalent to a Thouless or Brunswik ratio. It seems apparent from the statement quoted above that Sanders and Pacht are treating the mean constancy index of 88 for the psychotics as equivalent to Thouless' ratio, and hence they consider their findings similar to those of Thouless. It will be shown presently, however, that their index of 88 actually indicates overconstancy rather than underconstancy. Hence, their results for the psychotics are opposite rather than similar to Thouless' findings for schizothymes. The fact that their results differ from those of Thouless is not so important as is the fact that their size constancy index signifies deviations from perfect constancy (100 for Sanders and Pacht and 1.0 for Brunswik or Thouless ratios) which are opposite in direction from the usual size constancy index. For example, a Brunswik or Thouless ratio of less than 1.0 indicates underconstancy, i.e., a judgment which falls somewhere between retinal proportions and real object proportions. An index greater than 1.0 indicates overconstancy, i.e., a judgment which goes beyond real object proportions in a direction opposite from retinal proportions.

---

[1] The author wishes to thank Drs. M. Ray Denny and Paul D. Greenberg for their assistance and suggestions. However, the author assumes sole responsibility for the content of the note.

[2] Note that the terms *standard* and *comparison* have been substituted for Sanders and Pacht's *stimulus squares* and *fixed standard series*, respectively, in order to conform to the terminology generally found in the literature.

Sanders and Pacht point out that the psychotic group's size judgments go beyond real object proportions in a direction away from retinal proportions. Hence, this would indicate that their mean constancy index of 88 indicates not underconstancy but overconstancy. Furthermore, this becomes even more apparent when we consider the relative sizes of the matches for the standard at the two distances. Underconstancy would be indicated when a subject's match for the standard at a distance of four meters is smaller than his match at two meters since the size of the retinal image of an object decreases with distance. However, for these matches Sanders and Pacht's formula would have given an index of more than 100. When the subject's judgment for the far standard is larger than that for the near standard (overconstancy), the resultant Sanders and Pacht index would be less than 100. This inverse relation to the Brunswik or Thouless ratios becomes most apparent when one examines a hypothetical situation wherein our subject's judgments conform strictly to retinal proportions. With Sanders and Pacht's experimental arrangement, when the standard is 50 mm. in size, our subject's judgments would be 125 mm. for the near standard and 62.5 mm. for the far standard. Such a situation would result in Brunswik and Thouless ratios of zero, whereas Sanders and Pacht's index would be 150.

If one does not take into account the discrepancy between Sanders and Pacht's constancy index and Brunswik or Thouless ratios, much of their discussion seems confusing and appears to contradict their results. For example, they point out that their data indicate a breakdown in constancy for schizophrenics opposite in direction from that predicted by Bruner [1]; i.e., the schizophrenics' judgments, rather than conforming more with retinal proportions, go beyond real object proportions in a direction opposite from retinal proportions. This statement seems to contradict their report of a mean constancy index of 88 for the psychotic group when this index is accepted as equivalent to a Brunswik or Thouless ratio. However, when the fact that their index indicates deviations from perfect constancy in a direction opposite from Brunswik or Thouless ratios is taken into account, then the seeming contradiction disappears.

It is recognized that the usual methods of computation would not have given a single size-constancy index for the standard at two positions, but rather, two indices, one for each position of the standard. However, if the authors desired a single index, they could have obtained indices comparable to the usual ones simply by altering their equation to read as follows:

$$\text{Size-constancy index} = \frac{\text{Size}_{\text{far}} - \text{Size}_{\text{near}}}{\text{Size}_{\text{far}}}.$$

Just as for their equation, a constant of one should be added.

One wonders why, in the interest of consistency with previous research in this area, the authors chose an index not equivalent to those commonly used.

Received September 18, 1953.

References

1. Bruner, J. S. Personality dynamics and perceiving. In R. Blake, & G. Ramsey (Eds.), *Perception — an approach to personality.* New York: Ronald, 1951. Pp. 121-145.
2. Sanders, R., & Pacht, A. R. Perceptual size constancy of known clinical groups. *J. consult. Psychol.*, 1952, **16**, 440-444.
3. Thouless, R. H. Individual differences in phenomenal regression. *Brit. J. Psychol.*, 1932, **22**, 216-241.

NEW BOOKS AND TESTS

Books

Bach, George R. *Intensive group psychotherapy.* New York: Ronald, 1954. Pp. xi + 446. $6.00.

Intensive Group Psychotherapy attempts a marriage of group therapy and Kurt Lewin's group dynamics. Based on two continuous intensive therapy groups and liberally interspersed with protocols and brief examples from the therapy sessions, the volume is rich in its formulations of the many factors involved in the group therapy process. Its weakness is that, if anything, it is overorganized, possibly giving the student in group therapy the feeling that therapy can be done by rule of thumb methods. There are, for example, 4 personality criteria for exclusion, no more, no less, 16 communication themes, 7 therapeutic functions in role playing, 7 phases of "theragnostic work," 18 do's and 10 don'ts, 7 developmental phases, 7 communicational levels, 7 instigators of defensive resistance, and 8 factors of group dynamics. In spite of the mechanistic organization, this volume is one of the most intriguing works on group therapy to appear in some time. It is a brave attempt to bring order out of chaos, even if it is rigidly structural. If one accepts the author's axioms, the conclusions follow logically.—*M.K.*

Christie, Richard, & Jahoda, Marie. (Eds.) *Studies in the scope and method of "The authoritarian personality."* Glencoe, Ill.: Free Press, 1954. Pp. 279. $4.50.

In this unusual volume, seven authors make a retrospective evaluation of *The Authoritarian Personality* (1950) by examining some of its concepts, its methods, and the subsequent studies it inspired. Shils and Laswell contribute essays that deal with the study's structure of concepts and theory. In perhaps the most telling chapter, Hyman and Sheatsley re-examine problems of method with respect to sampling, measuring instruments, and methods of analysis. Christie reviews the newer research studies. One salient point that emerges both from the critique and from subsequent research is the need to control education and social class in studies of authoritarianism. Some of the attributes and correlates of the trait now seem to be identified as social class differences. In a concluding section, Frenkel-Brunswik reviews the study from the viewpoint of an original contributor, and reports new data from an investigation of children's prejudices. This critical volume gives a global impression that *The Authoritarian Personality* possessed many faults of method and hence of conclusions, but that it opened to investigation a significant area hitherto dominated by opinion and controversy.—*L. F. S.*

Eissler, Ruth S., *et al.* (Eds.) *The psychoanalytic study of the child.* Vol. VIII. New York: International Universities Press, 1953. Pp. 412. $7.50.

This eighth volume of a now well-established series is up to the excellent standard of previous volumes. The nonanalytic child psychologist should not be frightened by its theoretical orientation, as the case material is extensive and worth reading in its own right. The result is a mature and stimulating symposium.—*W. A. H.*

Eysenck, H. J. *Uses and abuses of psychology.* London & Baltimore: Penguin Books, 1953. Pp. 318 (paper). 65¢.

In this smoothly and interestingly written nontechnical book, Eysenck discusses intelligence testing, vocational psychology, abnormal behavior, and social attitudes. As might be predicted from his other writings, the author emphasizes the findings from research and takes a dim view of intuitive appraisals, psychoanalytic theory, and psychotherapy in general.—*L. F. S.*

Hartley, Eugene L. (Ed.) *Doubleday papers in psychology.* Garden City, N. Y.: Doubleday, 1954.

Leuba, Clarence. *The natural man.* Doubleday Papers in Psychology, No. 1, 1954. Pp. x + 70 (paper). 95¢.

English, Horace B. *The historical roots of learning theory.* Doubleday Papers in Psychology, No. 2, 1954. Pp. viii + 21 (paper). 65¢.

Frank, Lawrence K. *Feelings and emotions.* Doubleday Papers in Psychology, No. 3, 1954. Pp. viii + 38 (paper). 85¢.

These three pamphlets are the first of an announced series of fourteen papers on psychological topics. As judged from the initial sample, the pamphlets are clear, elementary, interesting, and at-

---

Note.—The reviews were prepared by the Editor and the Associate Editors, who may be identified by their initials.

tractive. An assembly of appropriate topics might provide text or reference material for a course in introductory psychology or for extension courses. The pamphlets would not be suitable for advanced or graduate instruction. Leuba's *The Natural Man,* an excellently integrated presentation of studies of chimpanzees and of nonliterate cultures, provides a novel introduction to the problems of psychology. English, in the second pamphlet, reviews theories of learning from Descartes and Hobbes to E. L. Thorndike. Frank's essay, *Feelings and Emotions,* is somewhat more diffuse and "popular" than the other two papers. It cites no references, ignores many issues that most psychologists would regard as relevant to its topic, and digresses into the areas of child development and personal adjustment.—*L. F. S.*

Ingham, Harrington V., & Love, Leonore R. *The process of psychotherapy.* New York: McGraw-Hill, 1954. Pp. ix + 270. $5.00.

This is a description of the way in which a psychotherapist works, with no attempt to introduce dynamics of personality except when necessary to understand treatment procedure. The noncontroversial conditions preliminary to, and occurring during, therapy are presented somewhat succinctly and simply. Two cases are included. One is a verbatim reproduction of treatment covering a single hour; the second describes events in 36 psychotherapeutic interviews. The absence of even a single reference to other books or periodicals is striking. There is no indication as to whom the book is addressed, but the authors might have indicated more emphatically the dependence of the technique upon an understanding of personality theory.—*F. McK.*

Libo, Lester M. *Measuring group cohesiveness.* Ann Arbor, Mich.: Univer. of Michigan Press, 1953. Pp. ix + 111 (paper). $2.00.

This research bulletin contains a discussion of the concept of group cohesiveness, a report of research on its measurement, and a new projective test for appraising attraction to group. The test shows excellent scoring reliability and significant relationships to behavioral measures of group cohesiveness but is presented for experimental use, not as a finished instrument for prediction.—*L. F. S.*

Klopfer, Bruno, Ainsworth, Mary D., Klopfer, Walter G., & Holt, Robert R. *Developments in the Rorschach technique.* Vol. I. *Technique and theory.* Yonkers, N. Y.: World Book Co., 1954. Pp. x + 726. $6.50.

Replacing the widely used *The Rorschach Technique,* by Klopfer and Kelley, this fat volume is at once a manual of administration, scoring, and interpretation, and a theoretical monograph that attempts to bring the Rorschach test within the purview of contemporary theories of personality. Probably the great majority of clinical psychologists will find something of interest and value here, a situation which attests at once to the book's comprehensiveness and to the depths to which the Rorschach has penetrated the clinical culture. One anticipates that the same kind of comment will be in order when a succeeding volume, *Fields of Application,* is published. In its role of manual, *Technique and Theory* represents a considerable advance over the older Klopfer and Kelley. There is much more detailed attention to scoring problems and to the clarification of scoring principles. The matter of interpretation is discussed with thoroughness, and with a greater degree of awareness than before that there are important validation problems still unsolved. The couching of interpretative assertions in the form of hypotheses may represent a decided advance in thinking about the place of the Rorschach in the clinician's tool kit, at once warning against dogmatism and providing an impetus for serious research on the instrument's actual utility. This part of the book is informative, teachable, and most welcome. The three chapters comprising the "theoretical monograph" are stimulating, if a bit uneven. Holt's chapter on the implications of contemporary personality theories for a Rorschach rationale is exceptionally able and constitutes the most serious and sophisticated effort to date to relate the test to broad and potentially testable issues in theoretical psychology. One senses a degree of defensiveness in Dr. Ainsworth's discussion of validation problems, but she has done a fair, intelligent, and thorough evaluation of the voluminous literature, and her chapter will be invaluable as a source of information. Bruno Klopfer's concern with Rorschach hypotheses and ego psychology is essentially speculative and given to too certain interpretative assertions. ("The earliest communicated response to shading shows a mixture of *cF* and *Fc*. This represents a considerably differentiated need for affection and belongingness" p. 590.) But read with an eye both to clinically useful insights and to the formation of hypotheses about Rorschach and about personality functioning, this vigorous chapter is challenging and evocative. There can be no gainsaying the importance of the book at this particular stage of clinical psychology's development. The judgment of history will depend on the quality and outcome of the research that it should stimulate.—*E.J.S.*

Paterson, Donald G., Gerken, C. d'A., & Hahn, Milton E. *Revised Minnesota Occupational Rating Scales.* Minneapolis: Univer. of Minnesota Press, 1953. Pp. ix + 85 (paper). $2.00.

This 85-page pamphlet is a second revision of the Minnesota Occupational Rating Scales, first offered in 1937. The scales consist of ratings, based on the pooled judgment of vocational psychologists, for seven abilities or aptitudes as these are required for each of 432 occupations. The abilities, each rated at four levels, are academic ability, mechanical ability, social intelligence, clerical ability, musical talent, artistic ability, and physical agility (new in this revision). The variation in the ratings from scale to scale yields an occupational pattern, e.g., ratings for eminent lawyer are A D A B D D D. These ratings are presented in a table in which jobs are arranged alphabetically, each preceded by an identifying number, and followed by the D.O.T. code number for jobs at every level of each ability, and a pattern classification that makes easy a consideration of occupational alternatives. Uses of the scales and the Individual Counseling Record are explained. By this system, both the counselee and the job are rated in the same terms.—*A. R.*

Tyler, Leona E. *The work of the counselor.* New York: Appleton-Century-Crofts, 1953. Pp. xi + 323. $3.00.

This is primarily a text for a first course in counseling. The author conceives counseling as "more a matter of what one perceives or comprehends, than what one says or does," and emphasizes this position most effectively in the context of discussing such practical matters as the use of records, tests, or occupational information. Because she admits that "counseling is still to a large extent an art," she writes largely in terms of wisdom based on "shared experiences," including her own. However, each chapter includes a research summary evaluating the relevant research evidence, and the final chapter is devoted to a summary of studies on the evaluation of the outcomes of counseling. Included is an up-to-date bibliography of 175 titles. Although designed as an introductory text, this little volume might well be read with profit by experienced counselors. It deals with basic issues and treats them in a manner that this reviewer found both challenging and sensible. The style of writing is superior.—*E. L. K.*

Witkin, H. A., Lewis, H. B., Hertzman, M., Machover, K., Meissner, P. Bretnall, & Wapner, S. *Personality through perception.* New York: Harper, 1954. Pp. xxvi + 571. $7.50

Witkin and his collaborators here report one of psychology's best achievements in the integration of theory, the laboratory, and the clinic. The major variable in their elaborate ten-year study is space orientation—the tendency of the subject to ". . . locate the upright according to the axes of the visual field or with reference to sensations from his own body. . . ." Tests of space orientation are studied in relation to one another, to other perceptual and motor tests, and to clinical appraisals, which include an interview, the Rorschach, the Figure-Drawing Test, and the TAT. Subjects are college students of both sexes, children from 8 to 17, and mental hospital patients. In the main, the data seem to confirm the hypotheses that "field dependence" in the perception of the vertical is a consistent variable, and that it has meaningful relationships to other features of the organization of personality. The studies of sex differences, and of the curvilinear relationships of certain perceptual variables with age from 8 years to adulthood, are especially penetrating. The methods for treating the data of the interviews and projective tests, achieving quantification without undue loss of clinical insights, may serve as models for future studies. Amid all these real merits, the work has one serious shortcoming. The interviews and projective devices were analyzed a posteriori by examining the performances of groups already identified as different in their space perceptions. The research with the personality tests is therefore exploratory and suffers from the classic error of discovering hypotheses and testing hypotheses with the same group of subjects. Cross validation is not entirely absent, but is weak and insufficiently stressed; procedures had to be modified for the children and for the patients in such a way as to leave indeterminate the extent to which the findings are due to the changed procedures, to the character of the samples, or to statistical shrinkage. The one genuine cross-validation group (p. 461) is small and incompletely reported. These faults can, and almost surely will, be remedied by further research studies. On the whole, the coordinated research program has great value, and the report should command the attention of all psychologists from the purest experimentalist to the most practice-centered clinician.—*L. F. S.*

Tests

Flanagan, John C. *Flanagan Aptitude Classification Tests.* Battery of 14 tests: Inspection, Coding, Memory, Precision, Assembly, Scales, Coordination, Judgment and Comprehension, Arithmetic, Patterns, Components, Tables, Mechanics, Expression. Grade 12-adult. 1 form (2 forms of 8 tests). 3 (5) — 35 (40) min. per test, (6) hrs. for battery. Test booklets ($3.20 per 25, each); aptitude classification sheet (70¢ per 25); personnel director's booklet, pp. 27 (50¢); counselor's booklet, pp. 35 (40¢); technical supplement, pp. 16 (40¢); student's booklet, pp. 20 (25¢); examiner's manual, pp. 27 (30¢); specimen sets ($3.00-$3.75). Chicago: Science Research Associates, 1953, 1954.

The impressive Flanagan test battery is a product of studies that began in 1946 and are still continuing. The 14 tests were devised to measure job elements based on analyses of the critical behaviors required in a number of occupations. For personnel selection and classification, selected batteries of tests may be administered, as recommended for the job or jobs to be filled. For vocational guidance, the entire six-hour battery may be used to obtain composite scores representing 30 occupations. The technical quality of the tests is excellent. Items have been invented ingeniously and are contained in well-planned booklets, most of which are arranged for easy self-scoring with carbon inserts. The reliabilities of the test elements range from about .55 to .85, but the reliabilities of the composite scores for occupations, each based on 2 to 7 subtests, are from .83 to .93. Test intercorrelations are generally low, with an average of .29 and only a few coefficients above .50. Norms, based on 1,563 Pittsburgh high school seniors, are expressed as stanines (single-digit standard scores) and as percentiles. Validity for occupational prediction is discussed in terms of a four-year follow-up study of high school students who entered seven occupations; raw correlations with criteria of vocational success range from .29 to .64. The five manuals provide a wealth of data cast in forms appropriate for the various types of users of the tests. In spite of the eight years of intensive work already given to these tests, they still have shortcomings. Some of the defects are easily apparent and are likely to be remedied by continued research, such as the needs for greater empirical validation with respect to specific occupations, and for more broadly based norms. Perhaps the most serious present fault comes from basing all norms on high school students. Thus, an average stanine score of 5 seems to imply average potential for every occupation, whether electrician or physicist, draftsman or engineer, salesperson or psychologist. In addition to the differential ability patterns of these occupations, *levels* must also be taken into account, or counseled students may be seriously misled. Expectancy tables based on the testing of persons in each occupation are needed for fully valid interpretation. Meanwhile, a student should be told, for example, that he might become an acceptable mechanic with a 4 on the mechanic scale, but to become a mathematician at all might require a 9 on the scale appropriate to that occupation. In spite of such remediable defects, the Flanagan tests seem now ready for use by personnel and counseling psychologists who are sufficiently well trained and discriminating. With further development, they are likely to become a very valuable tool.—*L. F. S.*

Books Received

Bauchard, Philippe. *The child audience.* Paris: UNESCO, 1953. Pp. 198 (paper). $2.00.

Bettelheim, Bruno. *Symbolic wounds.* Glencoe, Ill.: Free Press, 1954. Pp. 286. $4.75.

Feingold, Norman. *How to choose that career.* Cambridge, Mass.: Bellman Publishing Co., 1954. Pp. 52 (paper). $1.00.

Frank, Lawrence K., & Frank, Mary. *How to be a woman.* New York: Bobbs-Merrill, 1954. Pp. 144. $2.75.

Gillin, John, *et al. For a science of social man.* New York: Macmillan, 1954. Pp. vii + 289. $4.00.

Gray, J. Stanley. *Psychology applied to human affairs.* New York: McGraw-Hill, 1954. Pp. vii + 581. $6.00.

Ikin, A. Graham. *Life, faith and prayer.* New York: Oxford, 1954. Pp. 127. $2.50.

Lloyd-Jones, Esther, & Smith, Margaret Ruth. (Eds.) *Student personnel work as deeper teaching.* New York: Harper, 1954. Pp. xii + 361. $5.00.

Lorand, Sandor. (Ed.) *The yearbook of psychoanalysis.* New York: International Universities Press, 1954. Pp. 350. $7.50.

McCormick, Mary J. *Diagnostic casework in the Thomistic pattern.* New York: Columbia Univer. Press, 1954. Pp. xii + 239. $3.50.

Moloney, James Clark. *Understanding the Japanese mind.* New York: Philosophical Library, 1954. Pp. xix + 252. $3.50.

Pillay, A. P., & Ellis, Albert. (Eds.) *Sex, society and the individual.* Bombay, India (Whiteaway Bldg.): International Journal of Sexology, 1953. Pp. x + 448. $7.50.

Podolsky, Edward. (Ed.) *Music therapy.* New York: Philosophical Library, 1954. Pp. xii + 355. $6.00.

Sappenfield, Bert R. *Personality dynamics.* New York: Alfred A. Knopf, 1954. Pp. xiv + 412 + xvi. $5.50.

Simpson, George. *Man in society.* New York: Doubleday, 1954. Pp. vi + 90 (paper). 95¢.

Vedder, Clyde B. *The juvenile offender.* Garden City, N. Y.: Doubleday, 1954. Pp. xii + 510. $6.00.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

The Clinician as Judge: Implications of a Series of Studies

Roy M. Hamlin
Western Psychiatric Institute, University of Pittsburgh

In their most useful application, projective techniques are not concentrated estimates of single variables. They are not even multiscore tests. They permit extremely complex, unstructured behavior. The number and selection of variables are so little controlled that the behavior cannot conceivably be represented by measuring each ultimate factor separately. Statisticians like Cronbach [6] have attempted to suggest the mathematical complexity involved. Langmuir [16], the physicist, has stated that we must face the concept of "divergent phenomena," events so complexly determined that we cannot study them at all within the framework of classical science. He feels we must recognize the limits of science, and instead of science use common sense, judgment, and intuition.

At the present time, clinical psychologists use judgment. They describe and predict events or tendencies in the field of personality. For this purpose, they may use projective techniques as complex and particularly meaningful samples of individual behavior. The true value of these procedures is not represented in a single score or in a series of scores.

The reliability and validity of such techniques are not readily demonstrated. However, if psychologists can be persuaded to make limited judgments based on these techniques, the judgments can be tested for reliability and for validity. To a degree we can check the usefulness and accuracy of the procedures involved. In doing so, we follow the common procedure of science: from something extremely complex we tease out something simple enough to deal with objectively.

The studies of judgment with projective material have, so far, been somewhat disappointing. The reports on projective techniques in Kelly and Fiske's Michigan study of embryo psychologists [14], the Air Force studies using the Rorschach [20], the generally excellent California Institute of Child Welfare study of judgments of adjustment in a group of adolescent subjects [8], Kurtz's carefully fair use of the Rorschach in attempting to select insurance company personnel [15], and other similar studies have left the impression with both opponents and friends of projective techniques that such studies have so far shown zero or close to zero correlation between judgments based on projective techniques and any acceptably established outside criterion.

Purpose

The present discussion is limited chiefly to 10 studies, which used either so-called Rorschach signs, or judgments based on projective and clinical material. Five of these studies yielded positive results, and five failed to do so. The positive results were not entirely discouraging, and the negative results may be due to specific details of method. A consideration of these studies suggests that many factors contribute to positive or negative results. Chiefly for purposes of emphasis, however, one factor may be singled out for comment and stated in the form of a hypothesis.

Hypothesis: Positive and negative results in these studies are directly related to the simplicity or complexity of the material judged, and to the adequacy of the experimental conditions in enabling the clinician to derive judgments from the degree of complexity presented. *Subhypothesis 1*: In general, negative results may be expected when the units of material judged are too simple or atomistic. *Subhypothesis 2*:

In general, negative results may be expected when the units of material judged are too complex to be manageable within the conditions of the experiment.

It is probably apparent that the general thesis offered here is in some opposition, at least, to the use of so-called Rorschach signs and, on the other hand, to uncritical emphasis of global units or macrounits in clinical research. This opposition may be more apparent than real.

The Studies

The 10 studies considered were carried out by staff and graduate students associated with the University of Pittsburgh and with the Veterans Administration in the Pittsburgh area. Of these studies, four used Rorschach signs and six used judgments by clinicians. The studies have been reported elsewhere [1, 2, 4, 5, 7, 10, 11, 13, 18, 19, 21], and the two least available [13, 19] have been summarized by Newton [18]. Three other references to the series [3, 9, 12] are listed at the end of this article. Briefly, the 10 studies may be identified as follows:

1. Hamlin, Albee, and Leland [10] compared Muench's Rorschach signs for 71 subjects, divided into three groups of socially adjusted subjects, maladjusted subjects, and neuropsychiatric patients. The sign scores failed to distinguish between the groups.

2. Hamlin, Berger, and Cummings [11] compared Muench's Rorschach signs for 20 patients before and after six months or more of psychiatric therapy. Results were essentially negative though minimally in keeping with expectation.

3. Basic Rorschach Scores on the same 20 subjects [11] showed an average gain in adjustment score following therapy of 8 points, significant at the .01 level of confidence.

4. Newton and Goodman [19] compared Basic Rorschach Scores on 50 subjects with two different criteria of adjustment. Results were not significant.

5. Albee and Hamlin [1, 2] used a paired-comparison method to develop a scale of adjustment judged from drawings of a man and woman. Using this scale with 72 cases, judges rated normal subjects as significantly better adjusted than neurotics and as significantly better adjusted than schizophrenics. From these data, Bendig has derived a correlation of .47, significant at the .01 level of confidence between judgments and criterion groups.

6. Bialick [4, 5] used a paired-comparison method to develop a similar scale of intelligence judged only from five Rorschach *W*, or whole, responses for each subject. In details of method, this study is the most interesting of the series. For a small sample of 25 subjects, judgments and Wechsler-Bellevue IQ correlated .68, significant at the .01 level of confidence.

7. Cummings [7] developed 10 scales of adjustment, one for each Rorschach card. With 50 cases, the 10 combined judgments on each subject correlated .50 with an outside criterion of adjustment.

8. Newton [18] developed a scale of adjustment judged from total Rorschachs. With 50 cases, the Rorschach judgments failed to show a significant correlation with an outside criterion identical with Cummings', and with another outside criterion.

9. Hamlin and Newton [13] presented two Rorschachs from Newton's study to a seminar for "blind" analysis. In Newton's previous study, the two Rorschachs had been rated at exactly the same level of good adjustment, although one subject was known to be socially well adjusted and the other was psychotic and markedly maladjusted in nearly every respect. No other basis was considered in selecting the subjects. After careful study of the Rorschachs only, each of the eight members of the seminar rated as less well adjusted the subject known to be so, and the average group rating of the degree of difference in adjustment on Newton's seven-point scale was "reasonably accurate."

10. Whitmyre [21] had psychologists judge adjustment from drawings, and art students judge the same drawings for artistic quality. Each judge sorted and ranked 50 drawings as a single task, and an illustrated judgment scale was not used. The implications of the resulting high correlation (.78) between adjustment values and art values will not be considered in this paper. However, an incidental validity correlation between psychologists' judgments of cases and an outside criterion of adjustment was essentially negative.

Complexity of Units

The main thesis of this paper is related to the complexity of units judged. A rough continuum may be set up for the simplicity or complexity of the material:

1. Rorschach signs or other scored or judged single units that represent only small fragments of the projective material, mechanically selected.

2. Figure drawings or parts of the Rorschach selected on the basis of some rationale and retaining some degree of organization or configurational meaningfulness, some degree of complexity.

3. Total Rorschachs or units of similar complexity.

These three levels of units are illustrated by the studies of Newton [18], Cummings [7], Newton and Goodman [19], and Hamlin and Newton [13]. Newton used total Rorschachs, Cummings used single Rorschach cards, Newton and Goodman used Rorschach signs, and Hamlin and Newton presented two of Newton's 50 cases for blind analysis to a seminar where the protocols could be studied intensively and carefully. These four studies were essentially comparable in the judgment required, in the outside criterion, in the kind of subjects used, and in the projective technique considered. The four studies differed in the simplicity and complexity of the units judged, and in the comparable conditions provided for making the judgments.

Using Rorschach signs applied to Newton's 50 cases, Newton and Goodman found no significant relation with the relatively adequate outside criterion. The signs selected for this study were Buhler's Basic Rorschach Score.

With 10 different scales, one for each Rorschach card, and 10 separate judgments, one for each card, Cummings' resulting correlation of .50 was significant at the .01 level of confidence. The units judged were not too complex for the conditions of judgment.

However, when Newton used total Rorschachs as the basis of judgment, with entirely comparable subjects, judges, and outside criterion, no significant validity correlation was found. This result seems contrary to the idea that global judgments are more meaningful than judgments based on somewhat smaller units. However, the contradiction is not necessarily a real one. The one important factor that may be stressed is that the judging procedure broke down. The seven-point judgment scale was illustrated by a Rorschach at each point on the scale. These sample illustrations were too complex to be useful. No one judge rated more than 15 Rorschachs, and the judges spent ample time and effort on these. However, increasing complexity beyond a certain point probably does not lead to meaningful global understanding, except under ideal conditions or in a clinical situation where data can be cross-checked, discarded as not pertinent, or synthesized in a manner that eventually results in a simplified picture derived from complexity.

Stated in a crude way, the judgment process should be about as complex as the material judged and as the judgment required. For example, our seminar study of two of Newton's cases at least suggests that complex units of projective material, in this case total Rorschachs, can be used successfully in making judgments. These two Rorschachs had been rated at exactly the same level of adjustment by Newton's judges. However, under seminar conditions, every member of the seminar rated as better adjusted the subject so identified by the outside criterion, diagnosis, and over-all clinical picture.

In general, the results of all 10 studies may be interpreted as supporting the main thesis: Small, mechanically scored or judged units tend to give negative results. The likelihood of positive results is increased by larger units, by samples of behavior that are sufficient to represent some degree of organization, so that a clinical judgment—rather than a mechanical scoring without a rationale—is possible. However, the global unit, or macrounit, will give negative results unless the conditions for judging are complex, time-consuming, and approach actual clinical procedures. Briefly, the studies may be grouped in terms of unit size and judgment conditions:

Atomistic units. Of the four studies with Rorschach signs [10, 11, 19], three gave negative results. The one statistically significant result was probably due to an unusually good outside criterion.

Simple configurations. All three studies in which a judgment scale was established first and then applied to units smaller than global Rorschachs [1, 2, 4, 5, 7] gave results significant at the .01 level of confidence. Judgments of drawings correlated .47 with adjustment; judgments of single Rorschach cards correlated .50 with adjustment; and judgments of five *W* responses correlated .68 with intelligence.

Complex units. The three remaining studies [13, 18, 21] suggest the precautions that must be taken if the judge is presented with extensive, complex material in a single judgment task. Judgments of total Rorschachs gave negative validity results. Ranking 50 drawings as a single task, and without an illustrated judgment scale, gave negative validity results. Time-consuming, intensive analysis of two complete Rorschachs did permit correct judgments, but in these two cases only and under conditions that are impractical for many research purposes.

The California Child Welfare study of Rorschach judgments and adjustment [8] closely parallels one of our studies [18] in procedure and negative results. Other studies that have given negative results, and received wide attention, have used total Rorschachs or equally complex projective material, without providing ideal conditions for analyzing, digesting, and converting this material into the judgments required.

Other Considerations

Obviously, the complexity of units judged is not the only factor in determining positive or negative results. It is represented in the 10 studies considered, and it raises a more general question: Do we expect too much from clinical "experts"? A few other factors may be mentioned as of particular interest.

Degree of relationship expected. In planning a study in which clinicians' judgments of projective material are to be correlated against some second criterion, we should make a realistic estimate in advance of the degree of correlation that may reasonably be expected. Many psychologists tend to take a naive attitude toward clinicians and toward projective material. Clinicians *say* they can judge adjustment or introversive tendency from the Rorschach. Therefore, we expect their judgments to correlate at least .70 with any haphazard outside criterion: psychiatric diagnosis as a measure of adjustment, or a paper-and-pencil test of something that happens to be labelled introversion. We expect clinical "experts" to look at Rorschachs, for a relatively brief time, with material often dubiously reproduced, and then to pick good clinical psychologists with no idea what a good clinical psychologist is, or insurance sales managers with no rationale and no real clinical experience with these individuals. We tend to expect magic.

In an area where methodology is little developed, Cummings' correlation of .50 with a dubious outside criterion and an entity like adjustment, which has hardly been defined, is surprisingly high. Using only five Rorschach responses per subject, Bialick obtained a correlation of .68 with Wechsler IQ. Judgments on a mere fragment of the Rorschach correlate better with the Wechsler than does the CAVD, the Otis, or several other respectable tests of intelligence.

Definition and outside criterion. In regard to defining a variable and in regard to selecting and establishing an outside criterion, we should consider the known skills and limitations of the clinician. Bialick used intelligence as a variable. The judges were experienced clinicians with extensive knowledge of the meaning and measurement of intelligence. They were specifically acquainted with the operations involved in the test used as an outside criterion, the Wechsler. Furthermore, the outside criterion was a carefully constructed test of intelligence. It was not a sample of complex, real-life behavior only partially reflecting intelligence.

In using the variable "adjustment," we were on less firm ground. Adjustment was defined verbally and in terms of an outside criterion. The verbal definitions were sketchy, but in the later studies they tended more and more to include such phrases as "freedom from anxiety," "ability to handle anxiety," "contact with reality," "capacity for delusions or hallucinations." That is, we simply used other words to parallel our outside criterion groups: normals, neurotics, and psychotics. It might have been better to describe in precise and repetitive detail what the outside criterion was. The outside criterion closely paralleled the extensive and recent experience of the judges. The judges were already acquainted with the Rorschach performances of many neurotic and psychotic patients.

Material judged. Complexity of units has been discussed. One other point might be stressed in regard to material judged. With projective techniques especially, an adequate sample of material pertinent to the judgment required should be assured.

Bialick's study may be used for illustration. In the Rorschach, the blots and procedure do not require the subject to give responses within the framework of a specific variable to be judged. For this reason, a subject may not produce a sample of behavior pertinent to a specific judgment study. Bialick studied the quality of *W* responses as related to intelligence. Therefore, he took every Rorschach that had five or more *W* responses, but discarded any Rorschach that had fewer *W* responses. McReynolds [17] studied form level, and to do so had his subjects pass on responses that he selected in advance. The point is, for many research purposes, controls should be used to obtain an adequate sample of pertinent behavior. A well-integrated paranoid schizophrenic can give an innocuous Rorschach by default, and for clinical purposes his ability to do so is important. For some research purposes, however, we might do well to remember Rorschach's own comments on the need for control studies and control procedures. Ratios, correlational corrections, or "cannot say" categories for the judges are no substitute for an adequate sample of behavior.

Judges. In our expectations and plans for a judgment study, the clinician judge should be regarded as a subject, not a magician. His assets and limitations should be considered in decisions on complexity of task, sample of behavior presented, definition of variable, and selection of outside criterion. In addition to his background of general experience, he should have specific training in definition of variable, and should be given a trial run with correction of errors and opportunity to review his mistakes. Eventually, the basis of successful judgments should be thoroughly studied.

We might assume that the "expertness" of a competent clinician consists of a high degree of skill in descriptive prediction for subjects and variables with which he is extensively and recently acquainted in a somewhat limited setting. A good state hospital psychologist can probably say much that is valid and useful about state hospital patients. Put him in a correctional institution or a clinic for children and his predictive validity may be limited for a time. Use him as a judge in one of our studies and his predictive validity may be limited indeed. His skills may be readily converted or transplanted, but this process must be fostered and provided for. In this paper, complexity of units may illustrate the necessity for studying the clinician carefully in setting the judgment task.

Conclusions

In summary, the present paper has offered several suggestions in regard to studies of clinical judgment applied to projective techniques:

1. Such judgments are not like traditional test scores. In traditional tests, complex material is reduced to mechanical units that are counted. With projective techniques, the clinician may be asked to make an arbitrarily simplified judgment. Both methods serve to tease out something simple from something complex. But the methods are not the same.

2. A number of widely publicized studies have produced negative results in applying judgments to projective material. Usually, overly complex units of projective material have been employed.

3. In a few studies we have obtained positive results by using samples of projective behavior that were neither overly simple nor overly complex.

4. The positive results obtained in two of these studies have represented encouraging degrees of correlation. At the present stage, Cummings' correlation of .50 with a crude criterion of adjustment, and Bialick's correlation of .68 with Wechsler IQ may be considered respectable.

5. Incidental comment has been made on other methodological considerations. All comments emphasize careful study of the clinician in setting the judgment task.

Received August 17, 1953.

References

1. Albee, G. W., & Hamlin, R. M. An investigation of the reliability and validity of judgments

of adjustment inferred from drawings. *J. clin. Psychol.*, 1949, **5**, 396-398.

2. Albee, G. W., & Hamlin, R. M. Judgment of adjustment from drawings: the applicability of rating scale methods. *J. clin. Psychol.*, 1950, **6**, 363-365.
3. Albee, G. W., & Hamlin, R. M. The place of judgment in clinical research. *J. clin. Psychopathol.*, 1950, **11**, 174-177.
4. Bialick, I. A method of applying clinical judgment to problems of Rorschach validation: the quality of Rorschach whole responses and intelligence. Unpublished master's thesis, Univer. of Pittsburgh, 1950.
5. Bialick, I., & Hamlin, R. M. The clinician as judge: details of procedure in judging projective material. *J. consult. Psychol.*, 1954, **18**, 239-242.
6. Cronbach, L. J. Statistical methods for multiscore tests. *J. clin. Psychol.*, 1950, **6**, 21-25.
7. Cummings, S. T. The clinician as judge: judgments of adjustment inferred from Rorschach single-card performance. *J. consult. Psychol.*, 1954, **18**, 243-247.
8. Grant, Marguerite Q., Ives, Virginia, & Ranzoni, Jane H. Reliability and validity of judges' ratings of adjustment on the Rorschach. *Psychol. Monogr.*, 1952, **66**, No. 2. (Whole No. 334).
9. Hamlin, R. M., & Albee, G. W. Muench's tests before and after nondirective therapy: a control group for his subjects. *J. consult. Psychol.*, 1948, **12**, 412-416.
10. Hamlin, R. M., Albee, G. W., & Leland, E. M. Objective Rorschach "signs" for groups of normal, maladjusted and neuropsychiatric subjects. *J. consult. Psychol.*, 1950, **14**, 276-282.
11. Hamlin, R. M., Berger, B., & Cummings, S. T. Changes in adjustment following psychotherapy as reflected in Rorschach signs. In W. Wolff, & J. A. Precker (Eds.), *Success in psychotherapy*. New York: Grune & Stratton, 1952. Pp. 94-111.
12. Hamlin, R. M., & Kogan, W. S. Objectification in Rorschach interpretation. *J. Pers.*, 1948, **17**, 177-181.
13. Hamlin, R. M., & Newton, R. L. Comparison of a schizophrenic and a normal subject, both rated by clinicians as well-adjusted on the basis of "blind" Rorschachs. Paper read at Eastern Psychol. Ass., Atlantic City, N. J., March, 1952.
14. Kelly, E. L., & Fiske, D. W. The prediction of success in the VA training program in clinical psychology. *Amer. Psychologist,* 1950, **5**, 395-406.
15. Kurtz, A. K. A research test of the Rorschach test. *Personnel Psychol.*, 1948, **1**, 41-51.
16. Langmuir, I. Science, common sense, and decency. *Science*, 1943, **97**, 1-7.
17. McReynolds, P. Perception of Rorschach concepts as related to personality deviations. *J. abnorm. soc. Psychol.*, 1951, **46**, 131-141.
18. Newton, R. L. The clinician as judge: total Rorschachs and clinical case material. *J. consult. Psychol.*, 1954, **18**, 248-250.
19. Newton, R. L., & Goodman, H. W. A comparison of Basic Rorschach Scores with judgments of adjustment based on clinical case material. Paper read at Eastern Psychol. Ass., Atlantic City, N. J., March, 1952.
20. Sells, S. B. Description of the USAF School of Aviation Medicine research program on the development of psychiatric screening devices for selection of flying personnel. Paper read at Amer. Psychol. Ass., Washington, September, 1952.
21. Whitmyre, J. W. The significance of artistic excellence in the judgment of adjustment inferred from human figure drawings. *J. consult. Psychol.*, 1953, **17**, 421-424.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

The Clinician as Judge: Details of Procedure in Judging Projective Material[1, 2]

Irving Bialick
VA Hospital, Nashville, Tenn.

and Roy M. Hamlin
Western Psychiatric Institute, University of Pittsburgh

This paper reports an objective study indicating the correlation between judgments of intelligence based on Rorschach whole, or *W*, responses and Wechsler-Bellevue IQ. However, the study was planned essentially to illustrate and consider certain details of procedure in the use of judgment scales with projective material. In the section titled *Details of Method* we will highlight and comment briefly on specific considerations.

The hypothesis tested in the study was: experienced clinical psychologists can make valid judgments of intelligence based on five Rorschach *W* responses for each subject.

Subjects and Judges

The subjects of the study proper were 25 outpatients of the Pittsburgh VA Mental Hygiene Clinic. All were male, white veterans of World War II. The mean chronological age was 28 years, with a range from 20 to 39 years. All subjects had a psychiatric diagnosis of neurosis, with no reason to suspect the presence of schizophrenic reaction or of organic intracranial involvement. The Rorschach and Wechsler were administered as part of the routine psychological evaluation. No subject was included who had not produced at least five *W* responses on the Rorschach. The subjects were further selected to provide a distribution of IQ's exactly proportionate to Wechsler's intelligence classification of the general population. No other basis of selection was used.

In addition to these 25 subjects, 10 other subjects were used in the construction of a judgment scale by the method of paired comparisons. Except in distribution of IQ, these 10 subjects were comparable in all respects to the 25 subjects. In terms of IQ, the 10 subjects were selected so that two subjects had Wechsler IQ's within three points of each of the following IQ levels: 80, 100, 115, 125, and 135.

The judges in the study were four VA staff psychologists, thoroughly trained and experienced with both Rorschach and Wechsler. For purposes of incidental comparison, four less experienced VA psychologists in training were used as a second group of judges. None of the judges had any prior knowledge of the cases used in the study.

The Judgment Scale

In the construction of the judgment scale, 10 subjects, 2 at each of five IQ levels, were used. From the Rorschach of each subject, five *W* responses were selected according to an arbitrary system. These responses, including inquiry, were typed on 5 × 8 cards. The 10 cards were then presented to four experienced judges in the 45 possible pairings. The pairs were presented in randomized order to two judges, and in reverse order to the other two judges. Each judge considered two cards at a time and decided which set of *W* responses had been given by the more intelligent subject.

From this paired-comparison method, the 10 subjects were arranged in order of intelli-

---

[1] This article is based in part on a master's thesis submitted by Dr. Bialick in partial fulfillment of the requirements for the M.S. degree, University of Pittsburgh, 1950.

[2] Published with the permission of the Chief Medical Director, Department of Medicine and Surgery, Veterans Administration, who assumes no responsibility for the opinions expressed or the conclusions drawn by the authors.

gence according to each judge's estimates. Rho correlations of validity and reliability, for these 10 subjects only, will be reported here rather than under *Results*. The rhos for individual judges against intelligence ranged from .71 to .87. The combined rankings of all four judges yielded a rho of .77, significant at the .01 level of confidence. The reliability rho for the ratings of two judges against two judges was .87, significant at the .01 level. These correlations were based on 10 cases, with two cases selected arbitrarily at each of five IQ levels.

Only 5 of the 10 subjects were used in illustrating the final judgment scale. At each of the five known IQ levels, that case was retained which had been judged correctly by the clinicians, and which best represented a distinct step interval. The five cases retained were ordered by the judges in complete agreement with the criterion of Wechsler IQ. Finally, a numerical scale value was calculated for each scale sample of five *W* responses [1, p. 217], and the subject's actual IQ was added to the scale card.

Procedure

In the study proper, each judge estimated the intelligence of 25 subjects. All that the judge saw for any one subject was five *W* responses selected from the subject's Rorschach according to an arbitrary system. Each sample of five *W* responses, including inquiry, was typed on a 5 × 8 card. The judge compared this card with the five scale sample cards, and placed the subject at one of the five points on the judgment scale.

Four experienced clinicians judged the 25 cases. For incidental comparison, four less experienced clinicians in training judged the same 25 cases.

The assigned scale values were used in calculating product-moment correlations for judgments based on Rorschach *W* responses with Wechsler IQ. The pooled judgments of two judges were correlated with those of the other two to obtain a reliability coefficient.

Results

The four experienced clinical psychologists made valid judgments of intelligence based on five Rorschach *W* responses for each subject. The correlations of their judgments against Wechsler IQ were .61, .62, .69, and .73. The mean correlation for any one judge, computed by Fisher's method of weighted z's [2, p. 363], was .68, with a standard error of .11. The reliability coefficient for two judges against two judges was .84.

The less experienced judges also made valid judgments of intelligence. For these less experienced judges, the four correlations were .45, .51, .54, and .59. It will be noted that the highest correlation for a less experienced judge was not as high as the lowest correlation for an experienced judge. However, the mean correlation for the less experienced judges, .54, was not significantly lower than the mean for experienced judges, .68. The reliability correlation for two less experienced judges against the other two was .69. All correlations were significant at the .01 level of confidence, except the lowest for a less experienced judge which was significant at the .04 level.

Because of the small *N* of 25, the validity rho of .77 for 10 cases obtained in constructing the judgment scale and reported in the section on *The Judgment Scale* is of interest. These 10 cases were additional cases not included in the 25 cases, and may be considered at least a suggestive replication.

Details of Method

The essential purpose of this small study was to illustrate and consider certain details of procedure in applying judgment scales to projective material. These details will be listed, with brief comment, in this section under the following headings: (*a*) rationale, (*b*) criterion, (*c*) material judged, (*d*) judges, (*e*) the judgment task, and (*f*) generalization from results.

Rationale. In general, studies such as this may be regarded as simplified checks on the projective approach and on specific claims for the significance of unstructured techniques. Eventually, a sufficient body of such studies should provide a sound basis for the clinical use of projective techniques. However, it seems necessary to emphasize repeatedly that the purpose is not to replace the whole projective approach with a set of scored scales which are a botched compromise, neither carefully thought out in construction as traditional measuring scales, nor retaining the advantages of the projective approach.

A second consideration in terms of general purpose relates to the question of validation. In this study, we report a correlation between judgments of unstructured behavior and scores on a highly structured test. The two samples of behavior represent approaches from two markedly different origins. However, we are not validating projective material in the sense of predicting future behavior in a real-life situation. A number of studies of projective material have reached for predictive goals that are too far in advance of our present knowledge, and have suggested a misleading answer to such a simple question as: Will a projective technique correlate as well with an intelligence test as another intelligence test will? Also, in this study, we were chiefly interested in details of the judging process. Details of procedure can be considered best in a context where the correlation is at least a point or so above zero.

Criterion. The outside criterion selected for this study was the Wechsler-Bellevue IQ. The purpose of the study determined this selection, not the significance of the variable as one of the most important reflected in projective procedures.

We chose an intelligence test because, in the field of single artificial variable measurement, intelligence was the variable our judges knew most about. By training and experience, the judges had developed a definition of intelligence that was markedly influenced by the actual operations involved in the Wechsler. Every judge had used the Wechsler extensively with patients similar to the subjects of the study. They had had informal clinical experience, with correction of errors, in estimating Wechsler IQ from Rorschach performance. When they made their judgments in the study, they knew that the outside criterion was the Wechsler. The judges were as familiar with the judgment called for, and with the outside criterion, as they were with the projective material judged. Finally, the outside criterion had an adequate degree of known reliability and validity.

Material judged. For each subject in this study, the material judged consisted of five *W* responses given by the subject in a routine Rorschach. We wanted a sample of material supported by clinical experience as related to intelligence, and a sample retaining some degree of configurational complexity. However, we also wanted the sample to be small, representative for each subject, and comparable for each subject.

Clinical experience over a period of time has tended to support *W* quality as related to intelligence. It is also probable that five *W* responses will retain much of the configurational flavor of the total Rorschach. At the same time, the sample for each subject was brief and could be typed on a 5 × 8 card. The judges were never overwhelmed by a unit of material that was so complex as to be bewildering. The selection of *W* responses made possible the elimination of any subject who did not happen to give an adequate sample of behavior pertinent to the judgment required. The material judged was also comparable for each subject. The problems that arise when, for example, the same judgment is made on a Rorschach with 100 responses and on a Rorschach with 15 responses were avoided.

Judges. In planning the present study, the specific training and experience of available judges were given careful consideration. These judges knew the projective technique, were thoroughly acquainted with the outside criterion in actual operation, and had worked with patients similar to those in the study. Their specific experience was both extensive and recent. This attention to specific and recent experience may be more important than considerations of general academic background and clinical training.

The judgment task. The present study was planned to make the judgment task as simple as possible, yet meaningful in terms of the judges' clinical experience. This consideration accounts for the small number of cases, and for the small sample of responses per subject. The judges could work rapidly. They had no feeling that their judgment energy was exhausted after the first two cases. At the same time, the *W* responses seemed to be units that were sufficiently complex to retain an adequate degree of configurational meaning. The judges felt that they had at least some basis for estimating intelligence.

The simple judgment scale was an aid rather than an encumbrance. Only five points were used on the scale. The five samples of behavior could be kept in mind. Each point was labeled with a known IQ as well as represented by a brief sample of *W* responses. Finally, misjudged samples had been eliminated from the scale. Since each clinician may make judgments on a different basis, falsely judged scale samples may be confusing.

Generalization. A final comment should be made in regard to generalizing from the results of such a study as this. Disregarding the small *N* and the desirability of replication, let us suppose that a given degree of correlation can be established between these judgments and Wechsler IQ. Are we then justified in saying that the Rorschach measures intelligence? or that other judgments made from the Rorschach are valid? These questions are obviously not answered. Possibly, however, with a sufficient body of studies, we may be able to set up general principles in regard to the conditions necessary and the limitations indicated for such judgments to be valid.

Summary

1. The study was planned with the specific purpose of illustrating and considering certain details of procedure in studies of judgment applied to unstructured material. Comment on some of these has been presented in the section *Details of Procedure.* The considerations discussed are related to the conditions necessary for valid judgments, and to the limitations indicated.

2. In the study, experienced clinical psychologists made reliable and valid judgments of intelligence based on five Rorschach *W* responses. The correlation with Wechsler-Bellevue IQ was .68; the reliability correlation, .84.

3. Less experienced psychologists in training also made reliable and valid judgments of intelligence on the same basis. Their judgments tended to be less reliable and valid, but not significantly so.

Received October 22, 1953.

References

1. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
2. Kelley, T. L. *Fundamentals of statistics.* Cambridge, Mass.: Harvard Univer. Press, 1947.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

The Clinician as Judge: Judgments of Adjustment from Rorschach Single-Card Performance[1]

S. Thomas Cummings
Veterans Administration[2] and University of Pittsburgh

A review of the literature in clinical psychology in the past two decades reveals an increasing interest in the process and results of clinical judgment. Among others, Murray [10] and Hunt [8] have long urged a more intensive focus on clinical judgment and its ingredients—Murray in calling attention to "psychology's forgotten instrument," the interpreter of test data, and Hunt in his consistent emphasis on the need to validate the clinician.

A strong impetus to investigation of clinical judgment stems from the rapid expansion of the field of projective techniques with their inherent validation problems. The data yielded by projective techniques, while for the most part qualitatively wealthy and clinically useful, have offered many vexing problems in categorization and quantification leading toward validation. Validation efforts based primarily on the use of isolated scoring abstractions, as for example the number of human movement responses in a Rorschach, do not represent an adequate replication of the essential operations involved in making inferences about personality from such techniques, nor have the validating efforts based on such abstractions been highly productive of significant relationships with personality variables. On the other hand, research with projective technique data which more closely approximates the operations of the clinician-in-action, i.e., emphasizes the use of clinical judgment, has often, although not uniformly, yielded significant relationships with such variables [1, 2, 3, 4, 5, 11, 12, 13]. Such variables as the complexity of the judgment unit and the definability and stability of criteria will limit research in this area, as Hamlin [7] has emphasized in his review of variables in clinical judgment research.

The experiment reported here evaluates the reliability and validity of judgments of adjustment inferred from single-card performance on the Rorschach test. It was designed to permit maximum use of clinical judgment in evaluating the test data, while presenting judges with a circumscribed judgment task, the performance of a subject on a single Rorschach card.

Method

This research was conducted in two phases: a scale construction phase and a scale evaluation phase. In the scale construction phase a ten-point adjustment scale was constructed for each of the 10 Rorschach cards. In the scale evaluation phase, the ability of judges using these scales to discriminate between the Rorschach single-card performances of subjects of varying adjustment status was evaluated. The reliability and validity figures reported here refer exclusively to the judgments of adjustment made in the scale evaluation phase.

Scale construction phase. In the first phase, a ten-point Rorschach scale of adjustment was constructed for each Rorschach card, based on the pooled preferences of six judges who judged the single-card Rorschach performances of the 10 scale construction subjects in terms of adjustment. The scale construction subjects were white male veterans of World War II who showed limited variation in age and educational characteristics, but extensive variation in adjustment status. In terms of nosology, the following gross categories were represented in the adjustment continuum: one normal (a nonvolunteer college undergraduate), five outpatient neurotics, two outpatient schizophrenics, and two hospitalized schizophrenics. The patients represented were chosen by the experimenter on the basis of case history information without reference to Rorschach data.

---

[1] Read at the meetings of the Eastern Psychological Association, Boston, April, 1953. Part of a dissertation submitted to the University of Pittsburgh, 1950. The author wishes to express his appreciation for the guidance received from his dissertation committee members, Drs. George W. Albee, Roy M. Hamlin, Gerald R. Pascal, and Carroll A. Whitmer.

[2] From the Veterans Administration Hospital, Aspinwall, Pa.

The six scale construction judges judged the single-card Rorschach performances of the 10 scale construction subjects using the method of paired comparisons. Typically, a judge was presented with 45 folders containing the pairing of each subject with every other subject on a Rorschach card and asked to indicate for each pair, "Which Rorschach performance indicates the better adjusted individual?" Wherever "adjustment" required definition in this research, as in instructions to judges, the following combination of characteristics was referred to them as defining adjustment: "ability to get along with other people, psychosexual maturity, freedom from anxiety and psychopathology, and perception of reality in accord with prevailing cultural norms."

The quantitative values assigned to the ten points on each of the adjustment scales were derived by the mean-proportion method for handling paired-comparison data, described by Guilford [6]. This method converts the proportion of total preferences awarded to each subject, out of the total possible number of preferences on a Rorschach card, into normal curve base line deviates. By this technique, considerable variability in the range and spacing of scale values is possible, based on the relative consistency of the judges in their preferences.

These procedures yielded a ten-point scale of adjustment for each Rorschach card, consisting of the ranked performances of the 10 subjects on each card and the scale values derived for them.

Scale evaluation phase. In this phase of the research, evaluation was made of the ability of judges using these adjustment scales to discriminate between the Rorschach single-card performances of members of five diagnostic groups, representing different levels of adjustment.

A sample of 50 scale evaluation subjects was assembled with 10 members in each of the five following groups: Group I, normals; Group II, nonhospitalized neurotics; Group III, hospitalized neurotics; Group IV, nonhospitalized schizophrenics; Group V, hospitalized schizophrenics. The normals in the sample were a group of nonvolunteer college undergraduates. The patients were gathered from two outpatient clinics and two hospitals by selecting consecutive cases that satisfied the sample's qualifying characteristics from the current files of these services. The general qualifying characteristics for inclusion in this sample were as follows: male, white, 20-55 age range, one-half year or more of college training, no suspicion of having intracranial pathology. Additional qualifying criteria were imposed for admission to subgroups, regarding length of hospitalization or nonhospitalization and manner in which diagnosis was reached [4].

It was recognized from the outset that definition of the independent variable, adjustment status, in terms of membership in these five diagnostic groups was a gross definition that would tend to attenuate the relationships evaluated [9]. However, such a gross criterion definition is frequently used in prevailing practice, and it was our opinion that if results suggestive of significance were obtained with such a criterion, further refinement of the criterion would lead to enhancement of the significance of these relationships.

The scale evaluation judgments of adjustment were completed by two judges, neither of whom had participated in the scale construction phase. These judges made their judgments of adjustment in the following fashion. The 10 single-card performances representing an adjustment scale for the Rorschach card in question were ranked before the judge with the instruction that these performances represented an adjustment continuum achieved by prior rating. He was then asked to familiarize himself with the adjustment levels represented by the scale members and, following this, to make his judgments of adjustment of the 50 scale evaluation subjects' single-card performances by allocating each to the appropriate point on the scale, in terms of its being most closely comparable with one of the scale members in adjustment level.

The ranks assigned to the 50 scale evaluation subjects by these judges using the scales were converted into predetermined scale values. Reliability coefficients were computed for the judgments of adjustment made on each card and for total series of judgments by product-moment correlation of these scale values. Spearman-Brown correction of these coefficients produced measures of these judges' joint reliability.

The validity of these judgments was evaluated primarily by determining the degrees of association between Rorschach adjustment scale values and adjustment group status, i.e., membership in one of the five diagnostic groups. In evaluating the degrees of association between these variables, the unbiased correlation ratio, epsilon, corrected for broad categories, was computed for individual cards and for the total series, i.e., the average score for all cards combined. A further refinement of the validity evaluation procedures which was made was evaluation of the significance of differences in total series mean adjustment score, using the t test. In all these computations, the average of the two scale values assigned by the two judges to a subject's Rorschach performance constituted his adjustment score for that card. "Total series mean adjustment score" either for individual subjects or for adjustment groups is the average of the scores achieved on all ten Rorschach cards.

Results and Discussion

Table 1 presents the reliability and validity coefficients for each Rorschach card and for the total series achieved by the scale evaluation judges.

Reliability. The judgments of adjustment made by the scale evaluation judges showed high degrees of reliability. Their reliability coefficient for the total Rorschach series was .89, a coefficient significant at the 1 per cent level. On individual cards their reliability coefficients ranged from .90 on Card III to .74 on Card V, all significant at the 1 per cent level of confidence. Despite the appreciable spread in values between the extremes of this series, statistical evaluation of the significance of intercard differences in reliability suggests that the judges showed little true differential reliability in making their judgments from card to card. Only comparison of the two highest coefficients on Cards III and IV with the lowest on Card V and Card III with the next lowest on Card I yielded ratios suggestive of significance at the 5 per cent level. With reference to consistency of judgment then, these judges showed high degrees of reliability in making their judgments of adjustment.

Table 1
Reliability and Validity Indices of
Scale Evaluation Judges' Judgments of Adjustment

	Reliability		Validity	
Card	r_c	p	ε	p
I	.78	.01	.29	
II	.86	.01	.46	.01
III	.90	.01	.50	.01
IV	.89	.01	.36	.05
V	.74	.01	.44	.01
VI	.86	.01	.41	.05
VII	.80	.01	.61	.01
VIII	.85	.01	.25	
IX	.82	.01	.40	.05
X	.86	.01	.43	.05
Total series	.89	.01	.50	.01

Validity. The validity of these judgments was first evaluated by correlating mean adjustment score and adjustment group status. The epsilon value between mean adjustment score from the total series and adjustment group status was .50, significant at the 1 per cent level. The epsilon values on individual cards ranged from .61 on Card VII to .25 on Card VIII, as reference to Table 1 will illustrate. The four highest coefficients on individual cards are significant at the 1 per cent level, the next four at the 5 per cent level, and the lowest two fall below the 5 per cent level. Thus, these judges' judgments of adjustment showed medium-order and statistically significant degrees of association with the criterion, diagnostic status.

Further evaluation of the ability of these judges to discriminate between the adjustment levels of these subjects was made by comparing differences in mean adjustment score between the five adjustment groups. Table 2 summarizes the sample's statistics for total series mean adjustment score. Inspection of these group means shows a consistent decline from Group I to Group V, albeit some of the differences in adjacent means are small. Evalua-

Table 2
Comparison of Adjustment Groups in Total Series Mean Adjustment Score

Statistic	Adjustment group				
	I	II	III	IV	V
N	10	10	10	10	10
Mean	1.33	1.30	1.13	1.05	.81
SD	.10	.15	.28	.15	.43
SE$_m$	.03	.05	.09	.05	.14

tion of the significance of differences in these group means is presented in Table 3. Here, 5 of a possible 10 significant differences result. The normals (Group I) are significantly higher than the bottom three groups and the nonhospitalized neurotics (Group II) significantly higher than the bottom two groups. Three other fairly sizable differences between means fall short of statistical significance because they involve relatively high variances within this setting of low degrees of freedom.

Table 3
Fisher *t* Ratios on Adjustment Group Differences in Total Series Mean Adjustment Score

Group	II	III	IV	V
I	.50	2.22*	4.67**	3.71**
II		1.70	3.57**	3.27**
III			.80	1.88
IV				1.60

* Significant at 5% level of confidence.
** Significant at 1% level of confidence.

In essence, then, the judges' judgments of adjustment of these 50 subjects' single-card Rorschach performances show strong positive relationships with the external criterion of adjustment used, diagnostic status. Admittedly, the size of the relationships found has little value for purposes of individual prediction but leaves little doubt as to the mutual implication of these variables, particularly in view of the vagueness of the adjustment concept and the grossness of the criterion.

The essentially positive results obtained in this research invite comparison with two similar studies, one by Grant, Ives, and Ranzoni [5] and the other by Newton [11], both of which related judgments of adjustment from the Rorschach to external criteria of adjustment and yielded negligible relationships. Both studies used total Rorschach protocols as the judgment unit. The former used a simple descriptive four-point scale of adjustment, and case-worker ratings of adjustment based on interviews as a validating criterion. The latter used a seven-point adjustment scale derived from the research reported here; ranked total Rorschachs illustrated the points on the scale; the external criterion was psychiatric judgments of case history information, which, incidentally, correlated highly (.79) with diagnostic status, the criterion used in the research reported here. Without evaluating other attenuating factors, which these authors review comprehensively, we would emphasize with Hamlin [7] that complexity of judgment unit, i.e., total Rorschachs, was a common factor in these studies which probably tended to diminish the obtained relationships. With 10 judgments of a less complex judgment unit, the single card, it would appear that a judge's opportunities for intensive coverage of his data are greatly enhanced. Even though the factor of intercard variability in performance is sacrificed by this method, the greater familiarity with one's data afforded by it through making 10 judgments of coherent parts of a whole would seem to more than compensate for this.

A related methodological factor in clinical judgment studies that has received little attention but probably also has an attenuating effect is similar to the factor of "esthetic fatigue" reported in esthetic judgment studies, something we could call "rater demoralization and fatigue." This appears to be primarily a function of complexity of judgment unit and size of total judgment task. Many judges who participate in clinical judgment studies report feelings of dissatisfaction and of being overwhelmed by having to retain and manipulate so many complex variables simultaneously and having to continue to do so for what is perceived as a mountainous mass of protocols. If such a factor exists, it is probable that it would have a further depressing effect on relationships obtained.

Finally, worthy of restatement, despite its obviousness, is the point that this and similar studies have limited implications for the use of the Rorschach or any test as a clinical tool. Inferences gained from a test in clinical use are typically evaluated within a context of other test and interview behavior data, frequently including health and historical information. Blind interpretation of tests is a worth-while training device, an interesting pastime, and a useful methodological feature of research on the bases and limits of clinical judgment. However, it provides crucial validation information chiefly in evaluating the assertions of single-test partisans regarding the use of their favorite techniques in isolation, not in evaluating the validity of these techniques as they are characteristically used in clinical practice.

Summary

A ten-point Rorschach scale of adjustment was constructed for each Rorschach card. Six judges used the method of paired comparisons in judging the single-card Rorschach performances of 10 subjects representing a broad range of adjustment. The 10 ranked single-card Rorschach performances plus the scale values derived for them comprised a scale of adjustment.

Two additional judges used these scales in evaluating the adjustment levels implicit in the Rorschach performances of 50 subjects of varying adjustment status (five groups of 10 subjects each of normals, nonhospitalized neurotics, hospitalized neurotics, nonhospitalized schizophrenics, and hospitalized schizophrenics).

These judges made judgments of adjustment that were highly reliable. The validity of their judgments as related to the criterion, diagnostic status, was for the most part of medium order, not permitting accurate prediction of diagnostic status from knowledge of mean adjustment score (average of judges' ratings) alone, but indicating statistically significant relationships between these variables. The indeterminate unreliability in the criterion was noted as probably attenuating the relationships obtained here.

Certain factors in clinical judgment research, such as complexity of judgment unit and "rater demoralization and fatigue," are briefly discussed.

Received November 20, 1953.

References

1. Albee, G. W., & Hamlin, R. M. An investigation of the reliability and validity of judgments of adjustment inferred from drawings. *J. clin. Psychol.*, 1949, **5**, 396-398.
2. Benjamin, J. D., & Ebaugh, F. G. The diagnostic validity of the Rorschach test. *Amer. J. Psychiat.*, 1938, **94**, 1163-1178.
3. Bialick, I. A method of applying clinical judgment to problems of Rorschach validation: the quality of Rorschach whole responses and intelligence. Unpublished master's thesis, Univer. of Pittsburgh, 1950.
4. Cummings, S. T. An investigation of the reliability and validity of judgments of adjustment inferred from Rorschach test performance. Unpublished doctor's dissertation, Univer. of Pittsburgh, 1950.
5. Grant, Marguerite Q., Ives, Virginia, & Ranzoni, Jane H. Reliability and validity of judges' ratings of adjustment on the Rorschach. *Psychol. Monogr.*, 1952, **66**, No. 2 (Whole No. 334).
6. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
7. Hamlin, R. M. The clinician as judge: implications of a series of studies. *J. consult. Psychol.*, 1954, **18**, 233-238.
8. Hunt, W. A. The future of diagnostic testing in clinical psychology. *J. clin. Psychol.*, 1946, **2**, 311-317.
9. Masserman, J. H. Diagnosis and prognosis in psychiatry. *J. ment. Sci.*, 1938, **84**, 893-946.
10. Murray, H. A. *Manual for the Thematic Apperception Test.* Cambridge, Mass.: Harvard Univer. Press, 1943.
11. Newton, R. L. The clinician as judge: total Rorschachs and clinical case material. *J. consult Psychol.*, 1954, **18**, 248-250.
12. Vernon, P. E. The significance of the Rorschach test. *Brit. J. Med. Psychol.*, 1935, **15**, 199-217.
13. Vernon, P. E. Recent work on the Rorschach test. *J. ment. Sci.*, 1935, **81**, 894-920.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

The Clinician as Judge: Total Rorschachs and Clinical Case Material[1]

Richard L. Newton
University of Pittsburgh

This research is an investigation of the value of a clinical technique (the Rorschach test) and the validity of clinical judgment as reflected in interpretations of total Rorschachs and clinical case material.

The study had three objectives. The first and primary objective was to ascertain to what degree psychologists' judgments of adjustment from Rorschach protocols agree with psychiatrists' judgments of adjustment from clinical material on the same subjects.

The second objective was to discover to what extent Rorschach judgments agree with psychologists' judgments of clinical case material. The third objective was to determine to what degree psychiatrists and psychologists agree in making judgments of clinical case material.

Method

Adjustment Scales

As a preliminary step, two quantitative scales of adjustment were constructed. Cummings [2] derived a psychiatric adjustment scale by having each of six qualified psychiatrists rank a group of 10 subjects in terms of over-all adjustment. The six psychiatrists were presented with abstracts compiled from the case histories of the subjects. The 10 subjects represented a range of adjustment from social adequacy to psychosis necessitating hospitalization. Normalized scale values were computed from mean ranks to construct a ten-point scale of adjustment. Cummings estimated the reliability of the psychiatric scale rankings for these 10 subjects by a split-judging group technique, which yielded the highly significant product-moment correlation of .95. In the present study, Cummings' scale was reduced to seven points to eliminate tied values and to produce more discrete spacing of the scale values. The elimination of three scale values did not narrow the range of adjustment represented. In the remaining seven subjects there were one socially adequate subject, three neurotics, one nonhospitalized schizophrenic, and two hospitalized schizophrenics. All were white male veterans of World War II.

A Rorschach adjustment scale was derived by having each of six qualified Rorschach workers rank the complete Rorschach protocols of these seven subjects in order of over-all adjustment. Normalized scale values were computed from mean ranks, producing a seven-point Rorschach adjustment scale. The reliability of the Rorschach rankings for the seven subjects was estimated by a split-judging group technique, which yielded the highly significant product-moment correlation of .94.

The Use of the Adjustment Scales

The scales for clinical histories and total Rorschachs were employed in obtaining judgments of the over-all adjustment of each of 50 experimental subjects made up of five groups of 10 subjects each. The five groups were socially adequate subjects, nonhospitalized neurotics, nonhospitalized schizophrenics, hospitalized neurotics, and hospitalized schizophrenics. The sample of 50 subjects was comparable in age, education, sex, color, and adjustment range to the seven subjects employed in constructing the adjustment scales.

Psychiatric adjustment scale. Each of 10 qualified psychiatrists was asked to judge abstracts of clinical material (basic data sheets) of 15 of the experimental subjects in terms of adjustment as represented by the range covered by the seven scale subjects. Each judge was presented with the basic data sheets of the seven scale subjects and used the relative scale placements of these subjects as a guide in placing each of the 15 experimental subjects at a point on the seven-point adjustment scale. Thus, three independent judgments of adjustment were obtained for each experimental subject; these made possible the computation of an estimate of the reliability of psychiatric judgments.

---

[1] This article is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree, University of Pittsburgh, 1952.

A group of 10 clinical psychologists also made judgments of adjustment from the basic data sheets of the experimental subjects in exactly the same manner as that employed by the 10 psychiatrists.

Rorschach adjustment scale. The same 10 psychologists, all skilled in the Rorschach, made judgments of adjustment for the experimental subjects from global Rorschachs. Each Rorschach judge was presented with the Rorschachs of the seven scale subjects and used the relative scale placements of these Rorschachs as a guide in placing each of 15 of the experimental Rorschachs at a point on the seven-point Rorschach adjustment scale. None of the Rorschachs judged by any Rorschach expert was obtained from subjects whose basic data sheets he had judged. Therefore, there was no possibility of a psychologist associating a subject's basic data sheet with that subject's Rorschach.

Results

The following results were obtained:

1. Psychiatrists' judgments of basic data sheets revealed a very high degree of reliability (.91).

2. Psychologists' judgments of Rorschachs revealed a high degree of reliability (.73).

3. Psychologists' judgments of basic data sheets revealed a very high degree of reliability (.94).

4. Psychologists' judgments of Rorschachs were compared with psychiatrists' judgments of basic data sheets to determine the degree of relationship or agreement between the judgments of the two disciplines when each discipline used a different instrument of evaluation. The resulting correlation coefficient (.09) was not significantly different from zero.

5. Psychologists' judgments of Rorschachs were compared with psychologists' judgments of basic data sheets to determine the degree of relationship between the two sets of judgments made by the same discipline with two different instruments of evaluation. The resulting correlation coefficient (.20) was not significantly different from zero.

6. Psychologists' judgments of basic data sheets were compared with psychiatrists' judgments of basic data sheets to determine the degree of relationship between the judgments of disciplines when both disciplines were using the same instrument of evaluation. The resulting correlation coefficient (.86) is significant.

7. Finally, psychologists' judgments of Rorschachs were compared with placement in the five nosological groups to determine to what degree the Rorschach judgments were associated with diagnostic categories. The resulting eta (.23) was not significantly different from zero.

Discussion

The question arises as to whether or not the criterion was a good one. The following evidence suggests that the criterion was good. Very high reliability resulted from the judgments of the criterion material by judges of both disciplines, thus showing agreement within each of the disciplines. The correlation (.86) between disciplines in judging the criterion material was very high also, showing general agreement among the "experts."

Newton and Goodman [4] compared the psychiatrists' judgments of the basic data sheets with placement in five diagnostic categories and obtained an eta of .79, significant at the 1 per cent level. These findings suggest that the criterion of psychiatric judgment is positively correlated with diagnostic classification. However, in the same study, Newton and Goodman found that psychiatrists' judgments of basic data sheets did not correlate significantly with Buhler Basic Rorschach Scores [1] of the 50 Rorschachs. Although their study constituted a different approach to Rorschach evaluation, the findings were consistent with those of the present study. The findings of Hamlin and Newton [3] suggest that concentrated study of limited data will produce positive results.

Summary

This research investigated the validity of clinical judgments of adjustment based on global interpretations of Rorschachs and on clinical case material. The method employed incorporated an adaptation of psychophysical scaling techniques wherein a Rorschach scale of adjustment was derived from psychologists' judgments of adjustment of Rorschach protocols and a psychiatric adjustment scale was derived from psychiatrists' judgments of adjustment based on clinical case material.

The use of these scales by psychologists and psychiatrists in judging the adjustment of 50 experimental subjects resulted in four basic findings:

1. High reliability was obtained in each of the three judging tasks: clinical case material by psychologists, by psychiatrists; Rorschachs by psychologists.

2. A significant relationship was obtained between psychologists' and psychiatrists' judgments of case material.

3. Rorschach judgments did not correlate significantly with psychologists' judgments of case material or with psychiatrists' judgments of case material.

4. Rorschach judgments did not correlate significantly with placements in five diagnostic categories.

Received November 18, 1953.

References

1. Buhler, Charlotte, Buhler, K., & Lefever, D. W. *Development of the basic Rorschach score with manual of directions.* Los Angeles, Calif.: Rorschach Standardization Studies, No. 1, 1948.
2. Cummings, S. T. The clinician as judge: judgments of adjustment from Rorschach single-card performance. *J. consult. Psychol.*, 1954, **18**, 243-247.
3. Hamlin, R. M., & Newton, R. L. Comparison of a schizophrenic and a normal subject, both rated by clinicians as well-adjusted on the basis of "blind" Rorschachs. Paper read at Eastern Psychol. Ass., Atlantic City, N. J., March, 1952.
4. Newton, R. L., & Goodman, H. W. A comparison of Basic Rorschach Scores with judgments of adjustment based on clinical case material. Paper read at Eastern Psychol. Ass., Atlantic City, N. J., March, 1952.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

Intra-Individual Consistency in Response to Tests of Varying Degrees of Ambiguity[1]

Arthur C. Carr
Creedmoor State Hospital and Adelphi College

An assumption implicit in all testing procedures is that the behavior elicited in the test is a representative sample of the behavior characteristic of the individual in other situations as well [29]. Projective techniques in particular lean heavily upon the assumption that all behavior manifestations, the seemingly superficial as well as the obvious, are expressive of the individual's personality [1, 17, 24].

Recent research, however, would seem to justify the caution of Bellak that "the projective hypothesis . . . has to be handled with considerable care" [1, p.5]. The literature is replete with findings which suggest that behavioral measures based on projective techniques may be inconsistent and unpredictable [3, 5, 7, 11, 16, 22, 25, 26, 27].

In clinical practice an assumption nevertheless is held that apparently unrelated intra-individual responses may be motivationally identical, i. e., may stem from the same causal structure. In light of this assumption, it is well to remember that tests merely record *behavior,* while clinical insights are arrived at through *inference* [17]. It follows that consistency or orderliness in personality must be found in what is inferred to lie behind the responses of the individual rather than in the varied overt aspects of his behavior.

The purpose of the present study was to investigate an aspect of the projective hypothesis dealing with the consistency of an individual's responses to projective tests that vary as to degree of ambiguity or structuredness. It seemed obvious from previous considerations that any meaningful consistency in behavioral expression must be found in what is "behind," or in what motivates, a given response rather than in some gross or surface feature of the response. Therefore, the diverse test responses were organized in terms of what could be inferred to be their underlying feeling tone, emotion, attitude, or motivating force. This quality was called the "inferred affect" of a response. In terms of relevance for personality dynamics, a classification of four fundamental qualities evolved: *hostility, positive feelings, anxiety,* and *dependency.*

The investigation studied the intra-individual consistency of responses elicited by the Sentence Completion Test (SCT), the Thematic Apperception Test (TAT), and the Rorschach test. The Forer [12] version of the SCT was chosen because of its systematic coverage of interpersonal relationships.

As is apparent in Table 1, each of the chosen tests contains stimuli that are assumed to elicit responses from which can be inferred the subject's attitude or feeling tone with respect to important human figures—mother, father, females, males, people, and persons in authority. These stimuli vary in ambiguity—from

---

[1] Reviewed in the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the author are a result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

Adapted from a dissertation submitted in partial fulfillment of the requirements for the Ph. D. degree, University of Chicago, 1952. The author would like to thank Dr. Donald W. Fiske for his kind assistance and understanding guidance throughout all phases of the research on which this article is based, as well as Drs. Hedda Bolgar and Julius Seeman, other committee members. The author is also grateful to those who served in varying capacities in scoring test data, particularly Mr. Austin Herschberger, who was especially generous with both time and clinical skills.

Table 1
Test Stimuli Pertaining to Various Interpersonal Figures

Figure	Forer SCT	TAT	Rorschach
Mother	20. His earliest memory of his mother was 35. Most mothers 60. My mother 76. When my mother came home, I 94. When he was with his mother, he felt 99. Mothers	Picture 6 BM (Elderly woman and young man)	Card VII "Mother" card
Father	9. His father always 16. Most fathers 33. My earliest memory of my father 70. When my father came home, I 88. Whenever he was with his father, he felt 96. Fathers	Picture 7 BM (Elderly man and younger man)	Card IV "Father" card
Females	21. When I meet a woman, I 26. Sisters 39. Most women 67. My first reaction to her was 71. As she spoke to him, he 81. Most women act as though	Picture 4 (Woman clutching man's shoulders)	Female percepts seen (Cards I-X)
Males	10. Men 13. My first reaction to him was 30. Most men 50. Brothers 59. While he was speaking to me, I 95. Most men act as though	Picture 9 BM (Men lying on grass)	Male percepts seen (Cards I-X)
People	15. His new neighbors were 38. I feel that people 49. A lot of people 66. When I meet people I generally feel 84. Most people are 93. Whenever he is introduced to people, he	Picture 2 (Farm scene)	Human percepts Animal percepts (Cards I-X)
Authority	23. When he met his boss he 36. Taking orders 73. People in authority are 77. When they told him what to do, he 91. Sometimes I feel that my boss	Picture 1 (Boy and violin)	Card IV

statements on the SCT directly referring to such people as "my mother," to stimuli of comparable but less obvious meaning, as the cloudlike Rorschach inkblot which has also been assumed by psychologists to represent the mother figure [21, 23]. On such an implied continuum, the TAT picture of the older woman and the young man (the "mother-son" card) would lie between the extremes. Such variations in stimuli are also present for other important figures. For example, the SCT includes items referring directly to such figures as mother, father, females, males, people, and authority figures, while the Rorschach and TAT provide means for assaying the same relationships through stimuli whose representational values are less immediately apparent.

The study was concerned primarily with the intra-individual consistency of responses elicited by varying stimulus representation of *each* interpersonal figure. It is to be stressed that these responses were classified in terms of their inferred affect, i.e., hostility, positive feelings, anxiety, and dependency.

A frequent assumption in clinical practice is that stimuli more similar to the object being considered will elicit tendencies closer to awareness than will stimuli that are less like the object. It is assumed that affect and impulses denied expression toward the real object, or any close representation of it, are nevertheless elicited by less "threatening" stimuli which, by virtue of certain qualities, still represent or "stand for" that object. Although Lazarus [18] has recently suggested the desirability of an even further extension of the ambiguity continuum than that represented in the usual battery of diagnostic tests, the tests used in this study represent a wide range on such a distance continuum. Therefore, it was believed that the resulting analysis would provide a comparison between the covert and the more overt aspects of personality functioning and reveal any disparity in such levels of function.

Method

The projective tests were administered to fifty male patients in a mental hygiene clinic. Formal diagnostic classifications undoubtedly would have revealed a range from mild neurosis to that of borderline psychosis. In spite of the differences in levels of adjustment, however, all subjects were functioning with some degree of success in everyday interpersonal relationships. No attempt was made to control the group for intellectual level, age, or diagnostic classification. It was believed that these factors within any one individual would be equally rather than differentially operative in all tests. The subjects ranged in age from 24 years to 41 years, with a mean of 32. Intelligence quotients (W-B) ranged from 84 to 147, with a mean of 111.

The tests were administered in the normal course of routine psychological testing. The entire 100 items of the SCT were administered to each subject. The pertinent items were later abstracted for purposes of this research. The Rorchach test was administered according to the methods suggested by Beck [2]. Seven cards were chosen from the TAT on the basis of the applicability of each card to assay the interpersonal relationship under consideration (Table 1). Three qualified psychologists chose these cards as most appropriate for eliciting affect toward each specified figure. Administration followed standard TAT procedure.

The inferred affect of all pertinent responses was independently scored by two qualified clinicians. Each test was independently scored, neither judge being aware of which response belonged to any subject.

It became obvious that the original intention of maintaining similar categories of inferred affect within and between tests could not be sustained, because all stimuli did not evoke productions whose qualities were characteristic of every anticipated category. Thus, while the SCT evoked a large proportion of positively toned responses, the only qualities inferred from Rorschach responses proved to be those of hostility, anxiety, and dependency, with a category of neutral or unclear responses. Similarly, the categories within a given test were not applicable to all the various figures. For example, no dependency category was inferred from TAT themes pertaining to females, males, people, or authority figures. The categories of inferred affect, therefore, were ultimately established in terms of the particular set of responses evoked from the particular group of subjects.

Rorschach scoring. After the Rorschach tests had been scored according to standard procedures established by Beck [2], every human and animal percept (*H, Hd, A,* and *Ad*), along with its accompanying inquiry, was extracted from the total record and entered on coded cards. The first responses to Cards IV (father card) and VII (mother card) were also entered on separate cards. This procedure was used, rather than attempting to arrive at scores based on over-all reaction to the card, as research had suggested the reliability of this method [4]. Judges rated each extracted response for inferred affect in terms of one of the following qualities: hostility, positive feelings, dependency, anxiety, and "other" (neutral or unclear). Through joint discussion of the criteria established by DeVos [9] and Elizur [10], they were acquainted with the general method for scoring. Although the numerous subcategories established by DeVos were not utilized, his quantitative approach to the scoring of Rorschach responses in terms of what he calls "affective symbolism" greatly facilitated the ordering of the data.

Of 662 responses so classified, judges agreed on 617, a percentage of agreement of 93.2. Percentage of agreement between judges was compared with that which would have been obtained on a chance basis. Critical ratios obtained on Rorschach scoring, as well as on other tests, indicated that on all comparisons the agreement obtained was significant at the .001 level. Differences in scoring were later resolved by discussion between the two judges. Where varying interpretations were not resolved (approximately 2 per cent of the total number), responses were necessarily placed in the "other" category.

TAT scoring. Prior to scoring the TAT, each story was extracted from the total record and coded so as to make later identification possible. It became obvious that a distinction had to be made between the theme's inferred affect and that of the characteristics overtly attributed to the test representative of the given interpersonal figure, since the qualities therein were sometimes dissimilar. For example, the theme of a TAT story involving the death of a kind, loving person would be scored *hostility* in contrast to the score based on the characteristics overtly attributed to the figure.

Overtly attributed characteristics (TAT-overt) were classified into two categories: positive feeling and hostility. This categorization was applicable to all productions except those elicited by Card 1, wherein an additional category was devised for those stories in which authority figures were not introduced.

The following criteria provided a basis for scoring TAT themes. These criteria pertain only to the quality inferred from the theme and are thus independent of the described characteristics of the person or persons in any TAT production:

Hostility

1. Pressure from partner or peer — prohibiting, compelling, punishing, quarreling, interfering.
2. Physical harm inflicted by peer or partner.
3. Physical harm inflicted to peer or partner.
4. Death or illness of peer or partner.
5. Disappointment in peer or partner.
6. Bad news to peer or partner.
7. Degradation or depreciation of peer or partner.
8. Departure — when they suggest conflicting goals, desire of hero to get away.

Dependency

1. Succor from peer or partner — seeking or receiving aid, advice, consolation, etc.
2. Nurture — bestowing aid, support, care.
3. Filial obligation — when duty and dependency predominate. If obligation implies hostility, this latter quality may predominate.
4. Departure — when departure is seen by hero as necessary to "becoming a man," etc.
5. Rebellion against authority, domination, dependency — when strivings for independence seem greater than hostile elements.

Anxiety

1. Concern over the future, outcome of situation.
2. Fear of what will happen, fear of the unknown.
3. Decision — where a difficult choice is posed.
4. Occupational concern.
5. Curiosity.
6. Moral struggle, illicit sex, guilt.
7. Departure — when departure is source of anxiety and grief to the hero.
8. Confession, admission of guilt.

In a total of 300 stories so rated, agreement between judges was 91 per cent on rating inferred effect of themes, and 94 per cent on overtly attributed characteristics of the figure being assayed.

SCT scoring. The original four categories of inferred affect were maintained in SCT scoring results—positive feelings, hostility, dependency, and anxiety. Neutral or unclear responses were categorized as "other." In general, scoring was done in terms of the obvious meaning of the completions, that is, taking completions at their "face value." Thus, "Most mothers *are good*" was scored *positive feelings* without questioning whether or not the real meaning was that "most mothers are good but mine is not," etc.

The category *positive feelings* pertained to all completions in which the primary attitude expressed toward the figure was that of acceptance, love, or security.

The category *dependency* included those attitudes which were primarily ones of dependency or immature compliance. Childishly toned responses, often pertaining to eating or being fed, or the expression of child-like feelings were classified in this category.

The category of *anxiety* was applied to all completions in which the primary attitude was that of fear, guilt, anxiety, or insecurity.

The category of *hostility* contained all those completions in which the primary attitude expressed toward the figure was that of avoidance, derogation, rejection, or pure hostility.

All other remaining responses which could not be fitted into the above categories were classed as *other*. Factual statements generally fell into this category.

A total of 35 sentence completions were scored for each subject. Agreement between judges was 1,594 out of 1,750 responses, or 91.1 per cent. Differences in scoring were resolved in discussion.

Statistical Treatment of Data

Ratings for each figure being assayed in each test were now available for every subject. Where the number of opportunities for such ratings varied (as with the number of female, male, human, and animal responses evoked by Rorschach stimuli), responses were equated by changing them into percentage of total responses [8]. Where the number was constant for all subjects (as with the completions for each figure on the SCT), occurrences of a given affect for each figure were simply totaled. Ratings for the total group of subjects were dichotomized in terms of the median point for number of occurrences (or presence and absence) of each affect category.

Each subject was now scored in terms of "high" and "low" placement in each affect category for each interpersonal figure. The tabulation compared "high" and "low" groups on each variable with the "high" and "low" group of every other variable.

Chi-square tests of independence were used to test the relationships between these variables [19]. Yates's correction for continuity was applied in all instances. One-tailed tests of significance were applied to all predicted relationships, i.e., the tabled probabilities for the chi-square distribution were halved because only positive relationships were to be considered as significant [15].

In the results presented, only these predicted relationships are dealt with. Ancillary findings are to form the basis of a separate report. The number of relationships predicted includes only those in which similar affect categories were maintained. Lack of a dependency category on certain figures assayed in the TAT, for example, eliminated the possibility of a consistency between it and dependency as elicited by the other tests' representations of the same figure. Since no similar category was present on which to have made a prediction of consistency, none was made. One could argue, however, that if real consistency prevailed, affect categories would necessarily have to be similar between all tests. Any dissimilarity between such categories should then be considered an exception to any basic consistency in personality expression. On the other hand, if a particular stimulus fails completely to elicit a certain affect category from a wide range of subjects, the evidence hardly should be interpreted as showing a lack of consistency within the subjects. It seems better to assume that the stimulus was structured in such a way (in spite of its relative unstructuredness) that cues

Table 2

Consistency of Inferred Affect for Predicted Relationships between Tests

Tests compared	Total number of relationships		
	Predicted	Significant at 5% level	Significant at 10% level
Rorschach and total TAT	22	6 ($p < .001$)	12 ($p < .001$)
Rorschach and TAT-theme	14	5 ($p < .001$)	9 ($p < .001$)
Rorschach and TAT-overt	8	1 (nonsig.)	3 ($p < .05$)
Rorschach and SCT	22	1 (nonsig.)	5 ($p < .05$)
Total TAT and SCT	27	9 ($p < .001$)	12 ($p < .001$)
TAT-theme and SCT	13	3 ($p < .05$)	4 ($p < .05$)
TAT-overt and SCT	14	6 ($p < .001$)	8 ($p < .001$)

for "appropriate" behavior were so obvious as to limit the nature of the responses.

Results

Of 22 predicted relationships between the Rorschach test and the TAT (Table 2), six were significant at the 5 cent level of confidence and twelve at the 10 per cent level. Marked contrast occurred, however, when affect variables based on the TAT themes were isolated, as five were significant at the 5 per cent level and nine were significant at the 10 per cent level. Comparison of Rorschach and TAT-overt variables reveals that of eight predicted relationships, only one was significant at the 5 per cent level of confidence and three at the 10 per cent level. The greater part of the consistency between the Rorschach test and the TAT appears to reside in Rorschach and TAT -theme inferences.

Of 22 predicted relationships between the Rorschach test and SCT, only one was significant at the 5 per cent level and five at the 10 per cent level. Relatively less consistency is manifested between these two tests than between any other two tests.

Of 27 predicted relationships between the TAT and the SCT, nine were significant at the 5 per cent level of confidence and twelve at the 10 per cent level. Of 13 predicted relationships between the SCT and TAT themes, three were significant at the 5 per cent level and four significant at the 10 per cent level. Of 14 predicted relationships between SCT and TAT-overt variables, six were significant at the 5 per cent level and eight at the 10 per cent level. The greater part of the consistency between the SCT and the TAT appears to reside in SCT and TAT-overt inferences.

To make the results between tests roughly comparable, the number of significant relationships was tested by the binomial expansion [28] to determine if that number was sufficiently great to indicate a nonchance occurrence. However, any interpretation of the resulting p values must necessarily be made with caution because of the lack of independence of the measures, the assumption on which such a test is essentially based.

Discussion

The extent of the relationships generally supports the conclusion that there is consistency in an individual's responses to tests of varying degrees of ambiguity.[2] That this consistency is found not in the overt manifestations of behavioral responses, but rather in an inferred explanatory construct related to motivational aspects of behavior, would seem consistent with the classic study by Frenkel-Brunswik [13].

It is obvious that the demonstrated consistency is far from perfect. To what degree this is dependent upon the somewhat arbitrary categorizing of test stimuli (particularly in regard to the TAT), on the inadequacy of the scoring system, or on factors more directly related to the subjects is not known.

The individuals chosen for this study comprised outpatients at a mental hygiene clinic. It seems likely that, as recently suggested by Hanfmann and Getzels [14], the degree of consistency between expressions elicited from different levels may be a function of personality integration. Greater consistency could then have been expected with a group of "normal" subjects, and less consistency in a group of frank psychotics. Only further research could verify this assumption, however.

In spite of the lack of perfect consistency, one point seems unequivocal: There is no support for the suggestion sometimes made that the opposite, or converse, of a given affect will appear at more overt levels as contrasted with that which exists at a deeper level. The trend is definitely in the direction of consistency rather than that of antithetical or diametrically opposed relationships, i. e., reaction formations.

---

[2] An analysis involving the intra-individual consistency *within* tests as reflected by the generalization of a given affect from mother to females, people, and authority figures, as well as from father to males, people, and authority figures, had revealed that the generalization within the various tests was only slightly above that expected on a chance basis. This finding would appear consistent with McClelland's statement that "while a set of generalized expectations about fathers and mothers or parent-child relationships probably do develop . . . they almost certainly do not *derive in any simple fashion* from the actual characteristics of the parents or the child's relationships to the parents . . . . There are many influences which enter into the formulation of such generalized conceptions" [20, p. 308].

This generalization, however, would hold only for whatever levels of functioning are represented by the chosen tests. This is an important stipulation as preliminary research had demonstrated no significant agreement between inferred attitudes to mother and father taken from SCT data and feelings expressed toward either parent on a rating scale. While apparently disqualifying SCT data as necessarily representative of overt manifestations, in the sense of being inwardly accessible and chosen for outward reporting, the finding would not be incompatible with the clinical impression that the SCT is a projective technique in the real sense of the word.

It is interesting to note that least consistency was revealed between the Rorschach and the SCT, tests representing the extremes in degree of stimulus ambiguity, while greatest consistency was found between each of these tests and the TAT. The findings perhaps could have been predicted on the basis of perceptual theory. Bruner has written: "Generally speaking, the less the structuredness or univocality of the stimulus input, the more striking the role of directive nonsensory factors will be in determining the perceptual organization" [6, p. 161].

Also related to the nature of the demonstrated relationships is the fact that the levels of inference involved, as well as the levels of personality function assumed to be tapped by the tests, are roughly equivalent, suggesting that these results are not fortuitous.

Thus, in spite of research evidence which would suggest that behavioral measures are inconsistent and unpredictable, it would appear that lawful relationships can be found in behavioral responses and that responses superficially dissimilar may nevertheless reflect a basic consistency and orderliness in perception and personality. That this can be demonstrated in the case of quantifiable variables which have meaning in terms of both projective and personality theory is evidence that a dynamic formulation of personality does not preclude the possibility of meaningful quantitative analyses of clinical data.

Summary

The intra-individual consistency of the inferred affect of responses to projective tests of varying degrees of ambiguity was studied. Variables were established in terms of given affect categories related to specified interpersonal figures. These categories were dichotomized about a median, and the statistical method used to determine the relationships between them was chi square.

Analysis of data demonstrated greatest consistency between the Rorschach test and the TAT, and between the TAT and the SCT. Least relative consistency was present between the Rorschach and SCT. The greater part of the consistency of the TAT with the Rorschach test lay in ratings based on TAT themes, while the greater part of the TAT's consistency with the SCT resided in ratings based on the overtly attributed characteristics of the varying interpersonal figures. The greatest consistency appeared among variables based on levels of inference as well as on those levels of personality functioning which approximate each other. While the obtained relationships showed far from perfect consistency, the extent of the demonstrated relationships justified the conclusion that there is a basic consistency and orderliness in perception and personality which may underlie the variability in behavioral responses.

Received December 1, 1953.

References

1. Abt, L. E., & Bellak, L. *Projective psychology.* New York: Knopf, 1950.
2. Beck, S. J. *Rorschach's test.* Vol. I. *Basic processes.* New York: Grune & Stratton, 1949.
3. Bender, I. E., & Hastorf, A. H. The perception of persons: forecasting another person's responses on three personality scales. *J. abnorm. soc. Psychol.,* 1950, **45,** 556-561.
4. Blake, R., & Wilson, G. Perceptual selectivity in Rorschach determinants as a function of depressive tendencies. *J. abnorm. soc. Psychol.,* 1950, **45,** 459-472.
5. Bray, D. W. The prediction of behavior from two attitude scales. *J. abnorm. soc. Psychol.,* 1950, **45,** 64-84.
6. Bruner, J. S. Perceptual theory and the Rorschach test. *J. Pers.,* 1948, **17,** 157-168.
7. Carp, Frances M. Psychological constriction in several projective tests. *J. consult. Psychol.,* 1950, **14,** 268-275.
8. Cronbach, L. J. Statistical methods applied to Rorschach scores: a review. *Psychol. Bull.,* 1949, **46,** 393-429.
9. DeVos, G. A quantitative approach to affect-

ive symbolism in Rorschach responses. *J. proj. Tech.*, 1952, **16**, 133-150.

10. Elizur, A. Content analysis of the Rorschach with regard to anxiety and hostility. *Rorschach Res. Exch.*, 1949, **13**, 247-284.
11. Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64**, No. 9 (Whole No. 315).
12. Forer, B. R. A structured sentence completion test. *J. proj. Tech.*, 1950, **14**, 15-30.
13. Frenkel-Brunswik, Else. Motivation and behavior. *Genet. Psychol. Monogr.*, 1942, **26**, 121-265.
14. Hanfmann, Eugenia, & Getzels, J. W. Studies of the sentence completion test. *J. proj. Tech.*, 1953, **17**, 280-294.
15. Jones, L. V. Tests of hypotheses: one-sided alternatives. *Psychol. Bull.*, 1952, **49**, 43-46.
16. Kelly, E. L., & Fiske, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: Univer. of Michigan Press, 1951.
17. Korner, Anneliese. Theoretical considerations concerning the scope and limitations of projective techniques. *J. abnorm. soc. Psychol.*, 1950, **45**, 619-627.
18. Lazarus, R. S. Ambiguity and nonambiguity in projective testing. *J. abnorm. soc. Psychol.*, 1953, **48**, 443-445.
19. Lewis, D., & Burke, C. J. The use and misuse of the chi-square test. *Psychol. Bull.*, 1949, **46**, 433-489.
20. McClelland, D. C. *Personality*, New York: William Sloane Associates, 1951.
21. Meer, B., & Singer, J. A note on the father and mother cards in the Rorschach inkblots. *J. consult. Psychol.*, 1950, **14**, 482-484.
22. Orange, A. J. Consistency of manner of apperception in tasks of varying structure. Unpublished doctor's dissertation, Univer. of Chicago, 1951.
23. Rosen, E. Symbolic meanings in the Rorschach cards: a statistical study. *J. clin. Psychol.*, 1951, **7**, 239-244.
24. Rosenzweig, S. Idiodynamics in personality theory with special reference to projective techniques. *Psychol. Rev.*, 1951, **58**, 213-223.
25. Samuels, H. The validity of personality-trait ratings based on projective techniques. *Psychol. Monogr.*, 1952, **66**, No. 5 (Whole No. 337).
26. Sanford, R. N., *et al.* Physique, personality, and scholarship. *Monogr. soc. Res. Child Develpm.*, 1943, No. 34.
27. Tupes, E. C. An evaluation of personality trait ratings obtained by unstructured assessment interviews. *Psychol. Monogr.*, 1950, **64**, No. 11 (Whole No. 317).
28. Wilkinson, B. A statistical consideration in psychological research. *Psychol. Bull.*, 1951, **48**, 156-158.
29. Wittenborn, J. R. The implications of certain assumptions in the use of the Thematic Apperception Test. *J. consult. Psychol.*, 1950, **14**, 216-225.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

Rorschach Concepts of Normality[1]

Ann Lawler Brockway, Goldine C. Gleser, and George A. Ulett
Washington University School of Medicine
St. Louis, Missouri

During the past decade the Rorschach test has been administered more and more frequently to "normal" individuals, either as military personnel or as members of control samples in experimental research. As a result of the increasing accumulation of data on such subjects, many of the widely held concepts regarding what constitutes a "normal" Rorschach protocol have failed to be substantiated. Papers by Linn [9], Brussel and Hitch [3], and Wischner, Rotter, and Gillman [12] have presented summaries of Rorschach findings among military personnnel, while Beck *et al.* [1] and Cass and McReynolds [4] have published data on samples of normal adults from varied occupations. Contributions such as these should cause, and to some degree have caused, clinicians to be more cautious in interpreting certain Rorschach characteristics as signs of pathology, and have made it necessary to revise many of the widely held tenets of what constitutes a normal protocol. It is becoming more evident that the perception of inkblots is affected by factors such as sex, race, age, intelligence, education, and set and that individual differences cannot be interpreted on the basis of personality variables alone. It appears necessary, therefore, that a careful accumulation of Rorschach data on representative samples of the total population stratified with regard to these factors be obtained, so that interpretations of pathology may be made on the basis of appropriate norms for each individual.

The point has been raised, however, particularly by Klopfer [8] and more recently by Beck *et al.* [1] that normality when regarded as a statistical average may differ widely from a qualitative definition of normality as meaning well balanced. This is admittedly true, but it does not mean that we can make interpretations which are not confirmed by the empirical data. A possible solution to this dilemma is to base norms on individuals who are not only engaged in everyday activities implying adequate adjustment but who have also been interviewed by a psychiatrist and considered to be relatively free from neurotic symptoms. Such a criterion, while far from perfect, would tend to yield a more homogeneous sample with regard to personality adjustment. The purpose of this paper is to make a beginning in this direction by summarizing Rorschach scores of a group of white subjects homogeneous with regard to sex, age, education, and set, and psychiatrically screened on the basis of personality adjustment. The scores for the well-adjusted members of this group are compared with those of the maladjusted, and also with scores obtained on a group of psychiatric patients, in order to examine the influence of personality adjustment on Rorschach percepts. Our results add further evidence that many of the rule-of-thumb "signs of normality" used by interpreters of the Rorschach are inadequate.

Procedure

A group of 151 men without history of psychiatric disease, and 40 male patients with a variety of psychiatric disorders in which manifest anxiety was considered the predominant symptom, were administered the Rorschach test by one female examiner and screened as to psychological adjustment by one male psychi-

---

[1] This report was completed under Air Force Contract Number 33(038)-13884. It is one of a series under R and D Project 21-37-002, Division of Psychiatric Screening of Flying Personnel, Department of Clinical Psychology, United States Air Force School of Aviation Medicine.

Table 1

Statistics on Age and Completed Education for 151 Normal Subjects and 40 Patients

Group	No. of cases	Age (years)		Education completed (grade)				
		Median	Range	Median	Distribution			
					8–11	12	13–15	16 and over
Adjusted	126	22	17–36	14	6	31	38	51
Maladjusted	25	22	18–33	15	0	7	6	12
Total normals	151	22	17–36	14	6	38	44	63
Patients	40	29	18–36	12	9	11	9	11

atrist. The Rorschach was administered second in a battery of psychological tests. The detailed procedure has been reported elsewhere [11]. All of the normal group were paid volunteers, oriented to the effect that the tests administered were being investigated as part of a proposed battery to be used in screening Air Force personnel. In this way it was felt that the set toward the test was relatively constant. The patients had been informed by their psychiatrists that the results of the tests would provide information useful in therapy. Thus their set differed somewhat from that of the "normals." All Rorschach protocols were scored as to determinants and content according to the Klopfer and Kelley system [8] and for location, form level, populars, and originals according to Hertz [6]. Additionals were counted as one-half in the tabulation of determinants and were given full value in counting the number of space responses. They were not included in the remaining scores.

Table 1 presents vital statistics (age and education) on the sample classified according to psychiatric appraisal. It may be noted that the patients tend to be slightly older than the normal subjects and include a somewhat smaller proportion of college graduates. The subjects for the normal group consisted of 38 students from a large metropolitan university, 39 students of a Protestant seminary, 36 members of the Air Force reserves, 15 regular Army personnel, 20 new Army recruits, and 3 miscellaneous. The individuals were engaged in occupations ranging from unskilled laborer to semiprofessional, but the majority of the subjects were either "white collar workers" or students. The patients were obtained from private practice or clinics, were for the most part gainfully employed, and also included some college students.

Results

Table 2 presents the 50th percentile scores in the major Rorschach scoring areas for the total normal group and the 50th percentile ranks found by Cass and McReynolds [4] on a group of 104 "normals" (median age, 34; median grades completed, 12). Their figures are shown only for those categories for which the scoring methods were the same. In comparing these values it should be noted that the statistics used by Cass and McReynolds are not quite identical with the exact median point used in this study. Even so, their findings are

Table 2

Median Scores for Various Rorschach Scoring Categories in a Normal Sample of Young Males Compared to 50th Percentile Ranks Obtained by Cass and McReynolds on Comparable Scoring Categories

Rorschach category	Present sample $N = 151$	Cass and McReynolds $N = 104$
R	26.2	22.5
M	2.0	2.0
FM	3.7	3.5
m (*m*, *mF*, *Fm*)	1.2	.5
FK	0.2	0.0
K (*K*, *KF*)	0.1	0.0
k (*k*, *kF*)	0.2	0.0
F	11.1	7.5
Fc	1.4	2.0
c (*c*, *cF*)	0.7	0.0
C′ (*c′*, *C′F*, *FC′*)	0.9	1.0
FC	0.9	1.5
CF	2.4	2.0
C (*C*, *Cn*, *C sym.*)	0.0	0.0
Σ*C*	2.8	3.0
Content		
H	2.4	2.5
Hd	1.0	0.5
A	9.1	9.0
Ad	1.4	1.0
P	5.6	5.5

very similar to those presented here, which is all the more remarkable since their sample included 46 females and a somewhat wider age and educational range than did the sample in our study. The major difference appears to be that they obtained a smaller median number of responses with an equivalent reduction in the median *F* response. They also obtained somewhat fewer responses determined by *c* and *m* and more categorized *FC*. It would be interesting to determine whether these differences could have been due to the inclusion of females in Cass's study. In both Cass's data and ours it may be noted that *FM* is greater than *M* and *CF* is greater than *FC*, although the reverse inequalities have usually been accepted as indicative of good normal adjustment.

Beck's results are not directly comparable to ours because of differences in scoring categories. Also, he presents his results in terms of means and standard deviations, despite the ex-

Table 3

Medians, 20th, and 80th Percentile Points on Various Rorschach Scoring Categories for a Normal Sample Classified as to Personality Adjustment and for a Sample of Psychiatric Patients

Rorschach Category	Adjusted normals ($N = 126$) Percentile scores			Maladjusted normals ($N = 25$) Percentile scores			Psychiatric patients ($N = 40$) Percentile scores		
	20th	50th	80th	20th	50th	80th	20th	50th	80th
R	15.9	26.6	39.9	14.7	24.9	30.5	15.5	22.2	33.8
Location									
W (*W*, *W*, *DW*)	4.7	8.4	13.6	6.2	9.3	13.5	4.0	8.0	13.8
D	5.8	12.1	22.3	4.3	11.0	18.5	3.5	10.5	19.5
dr	.8	2.9	7.5	.5	1.9	4.5	0	1.8	7.1
S	.9	2.9	5.9	.5	2.4	4.2	1.4	2.9	5.8
W%	16.7	33.2	58.7	21.8	41.8	70.5	16.3	30.5	65.5
D%	27.1	48.4	62.2	20.5	46.0	61.8	20.5	45.5	64.5
dr%	3.6	12.1	22.7	3.7	8.8	14.5	0	8.2	28.0
S%	3.6	10.1	18.8	.2	9.5	14.5	6.3	11.8	25.5
Determinants									
M	.9	2.0	3.4	.2	1.2	2.8	.2	1.1	3.1
FM	1.3	3.9	6.9	1.8	2.8	5.8	1.2	3.2	4.8
m (*m*, *mF*, *Fm*)	.1	1.0	2.7	.1	2.0	3.2	0	.9	1.8
FK	0	.2	1.2	0	.2	1.2	0	.1	1.0
K (*K*, *KF*)	0	.1	1.0	0	.1	1.0	0	.1	.9
k (*k*, *kF*)	0	.2	1.2	0	.4	1.2	0	.8	1.6
F	6.1	11.5	20.7	4.4	9.6	16.8	6.2	10.5	17.3
Fc	.2	1.4	3.1	.2	1.2	3.0	.1	1.0	2.8
c (*c*, *cF*)	0	.6	1.9	0	1.0	3.8	0	.2	1.6
C′ (*C′*, *C′F*, *FC′*)	0	.8	2.2	.2	1.2	2.2	.1	.9	2.1
FC	.1	1.0	2.2	0	.7	2.2	0	.8	1.8
CF	.4	2.2	4.2	1.9	3.2	5.2	.6	1.6	3.4
C (*C*, *Cn*, *Csym.*)	0	0	.2	0	.1	.2	0	.1	.25
ΣC	.9	2.6	4.6	1.8	3.9	5.8	.1	1.9	4.4
Content									
H	.9	2.6	4.5	.9	1.9	4.0	.6	1.5	3.1
Hd	0	1.1	3.6	0	.6	2.5	0	1.6	4.0
A	5.4	9.4	13.5	5.0	8.0	11.0	3.0	6.7	9.8
Ad	.2	1.9	2.0	0	.7	3.5	0	.3	3.5
At	.6	2.3	4.5	1.2	2.2	3.9	.3	2.1	5.8
Sex	0	0	.3	0	0	.3	0	.1	.4
Clouds	0	.1	.8	0	.1	1.0	0	.1	.2
Blood	0	0	.4	0	.2	.7	0	.1	.4
Ob., *Art*, *Arch.*	.4	2.4	5.7	.5	2.6	4.1	.7	2.3	4.5
Nat., *Geo.*	.8	3.3	6.5	1.2	3.4	8.8	.7	3.0	7.0
P	4.1	5.7	8.2	3.3	5.1	7.0	3.3	5.5	7.2
O	1.6	5.0	11.0	3.5	5.0	9.5	1.3	3.8	9.1

treme skewness of distribution of the scores in most categories. However, we might note that he also found mean *CF* greater than mean *FC* for his sample.

Table 3 presents the 20th, 50th, and 80th percentile scores in the major Rorschach scoring areas for the screened normals, the maladjusted normals, and the patients. It is believed that the "adjusted" group is a fairly representative sample of young adult white males of higher than average educational level and relatively good personality adjustment. Therefore, the middle 60% range of scores should approximate that which may be considered usual or typical for such a group.

From Table 3 it can be noted that the adjusted normals are slightly more productive of responses than either the maladjusted normals or the patients. Some interesting points of contrast between the adjusted and maladjusted normals of this sample are the latter's greater emphasis on whole responses at the expense of other location areas; their greater use of *c*, *m*, *C′*, and *CF* as determinants; and *blood, nature,* and *geography* as content categories. The adjusted normals on the other hand apparently use more *M*, *H*, and *A* than do the maladjusted subjects. Of all these categories however, only *M* and *m* yield a statistically significant difference at the optimal cutoff point.

The patients differ from the adjusted normals in their relatively greater emphasis on *S* responses. They also make greater use of *k*, *C*, and *C′* as determinants and *Hd* and *sex* as content particularly when these scores are considered relative to the total number of responses. On the other hand, it is particularly interesting to note that in some categories where differences are generally hypothesized, no differences are found for this sample. Thus, the adjusted controls utilized *K* (usually interpreted as free-floating anxiety) and *m* (interpreted as indicating tension and repression) to as great an extent as did the anxious patients.

Although trends were not necessarily to be expected in going from the adjusted normals through the maladjusted normals to the patients, since the maladjusted normals were so rated on a wide variety of symptoms whereas the patients were chosen primarily on the basis that they had manifest symptoms of anxiety, nevertheless some trends are apparent. It appears that as degree of maladjustment increases there is a diminution in productivity and also in the use of *M* as a determinant and *H* and *A* as content. On the other hand, there is an increase in the use of *k* and pure *C* as determinants. Unpublished research done at this laboratory has demonstrated that when *FC′*[2] is omitted from the *C′* category and cutting points are established with respect to number of *R*, the incidence of this determinant (*C′* and *C′F*) in protocols can be used to discriminate adjusted from the maladjusted subjects and the patients (*C′* + *C′F* less for adjusted subjects).

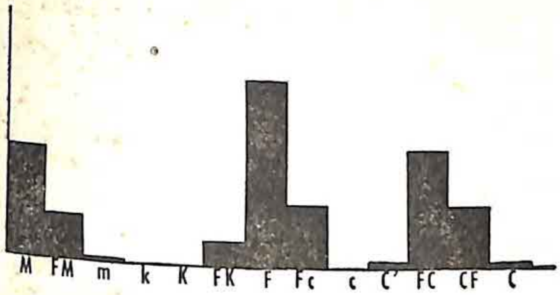
One of the most important facts to be obtained from Table 3 is that the range of scores for the middle 60% of the adjusted normals is often greater than is generally considered optimal by Rorschach workers (Klopfer, Brussel, and Hitch). This is particularly true of *W*% (17% – 59%), *dr*% (4% – 23%), *S*% (4% – 19%), *FM* (1.3 – 6.9), and *CF* (0.4 – 4.2). For other categories the middle range is shifted somewhat lower or higher than is postulated. It is suggested that optimal norms for Rorschach categories might better be developed by obtaining data on samples such as the present, in which psychiatric ratings are utilized to screen out the obviously maladjusted and then the middle 60% used to define the expected or optimal range.

Figure 1 presents the psychogram (using median scores) of the psychiatrically screened normals studied in this investigation along with a textbook illustration of the "normal" Rorschach psychogram [2, p. 128]. A comparison of these psychograms reveals that the screened "normals" used *F*, *FM*, *m*, *k*, *K*, *c*, and *C′* to a relatively greater extent than the hypothetical normal, with *M*, *FK*, *Fc*, *FC*, and *CF* being used to a lesser degree. It will also be observed that the ratios of *M* to *FM* and *FC* to *CF* differ in direction in the two psychograms, with *FM* and *CF* exceeding *M* and *FC* respectively in the screened normal group.

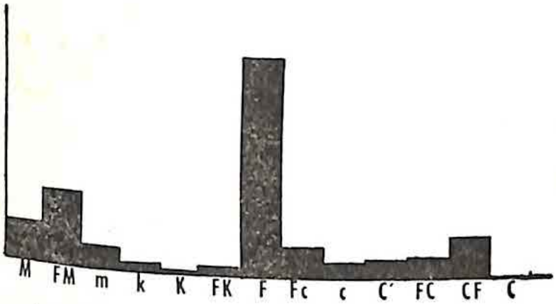
Table 4 shows the incidence of Davidson's [5] Rorschach signs of adjustment in the psychiatrically classified subgroups. Using com-

---

[2] Use of *FC′* as a determinant was demonstrated to be a sign of good adjustment.



HYPOTHETICAL NORMAL PSYCHOGRAM



ADJUSTED NORMAL SAMPLE N = 126

Fig. 1. Comparison of the hypothetical normal psychogram with that obtained on the present adjusted normal sample.

monly accepted Rorschach signs of adjustment but no outside criterion, Davidson reported the biserial correlation between the presence or absence of each sign and the total number of signs, thus relying on internal consistency as the test of validity. By comparing the percentage of each group showing a sign, it may be seen that a number of the signs discriminate in the direction opposite to that predicted by Davidson. Thus a larger percentage of the patients had $M \geq FM$ and $Dr+S \leq 10\%$ than did either group within the normal sample.[3] There was also a reversal in expected direction in the case of $Sum\ C > (Fc+c+C')$; $F\% \geq 20$ and ≤ 50; $\%R\ VIII\text{-}X \geq 40$ and ≤ 60; $A\% \leq 50$ and refusals, in that the percentage of maladjusted normals, and in some cases of patients, showing the signs was greater than of the adjusted. The signs that held up, though not always to a statistically significant degree even under a one-tailed hypothesis, were $M \geq 3$; $P \geq 4$ but $< .30\ R$; $R > 20$; $FC \geq 2$; no pure C; $FK + Fc \geq 2$; and $FC \geq CF$. For the latter sign it should be noted that it is necessary to exclude cases for which $FC = CF = 0$ in order to obtain the discrimination. Davidson did not specify this exception. Also for the P scores it was necessary to alter the criterion where R was less than or equal to 13. For these cases 4 P's were considered a sign of good adjustment.

Special note should be made of the incidence of shading and color shock in the three psychiatrically classified groups. Using the criteria proposed by Munroe [10], we found that only three of the subjects in the total normal group were *without* evidence of color shock (1.6% of the adjusted normals and 4% of the maladjusted normals) and 14 were *without* evidence of shading shock (7.9 and 16%). Though 100% of the patients showed some evidence of color shock, 10% of this group were without evidence of shading shock as compared to 7.9% of the adjusted normals. These findings are remarkable not so much as indicating a reversal in expected trend but as pointing out that what has been termed "shock" actually occurs to some extent in almost all Rorschach protocols. Thus these signs, at least when defined according to Munroe's criteria, cannot be used in detecting maladjustment. This criticism has also been posed in an article by Keehn [7], who makes a review of the literature with reference to color shock and concludes that the sign is of questionable value in the diagnosis of neuroticism.

That as great a proportion of the adjusted normals refused one or more cards as did the maladjusted is rather surprising, since card refusal has been accepted by most clinicians as pathological. The highest incidence of card rejection in our normal sample occurred in a group of 15 career military men, 13 of whom were sergeants and 2 commissioned officers. This subgroup had 5 members (33 1/3%) who rejected one or more cards. This finding is in agreement with the report of Linn [9] that "well-adjusted" military personnel frequently reject Rorschach cards. In our sample it was noted in particular that Cards V and/or VI were rejected more frequently by the "ad-

---

[3] The symbol $\geq$ means greater than or equal, and the symbol $\leq$ means less than or equal.

Table 4
Application of Davidson's Signs of Adjustment on Rorschach to Present Sample Classified According to Psychiatric Criterion

Rorschach sign of adjustment	Percentage of group in which sign was present			Probability of no difference (One-tailed hypothesis)	
	Adjusted normals	Maladjusted normals	Patients	Adjusted normals versus maladjusted normals	Adjusted normals versus patients
$M \geq FM \neq 0$	27.8	16.0	30.0	.11	Reversed†
$M \geq 3$	40.5	20.0	22.5	.03*	.02*
$\Sigma C > Fc + c + C'$	52.4	64.0	45.0	Reversed	.21
$20 \leq F\% \leq 50$	59.5	64.0	57.5	Reversed	—
$Dr + S \leq 10\%$	37.3	48.0	55.0	Reversed	Reversed
$P \geq 4 < .30\ R$ (Except when $R \leq 13$, $P = 4$)	69.0	60.0	57.5	.19	.09
$R > 20$	68.3	64.0	55.0	—	.06
$FC \geq CF \neq 0$	36.5	20.0	30.0	.06	.22
$FC \geq 2$	31.0	24.0	20.0	.24	.09
No pure C	86.5	76.0	80.0	.09	.16
$\%\ R$ VIII-X $40 \leq R\% \leq 60$	34.1	36.0	20.0	Reversed	.05*
$FK + FC \geq 2$	62.7	60.0	47.5	—	.04*
$W{:}M > 1{:}1$ and $< 3{:}1$	21.4	12.0	20.0	.15	—
$A\% \leq 50\%$	67.5	80.0	72.5	Reversed	Reversed
No color shock†	1.6	4.0	0	Reversed	—
No shading shock‡	7.9	16.0	10.0	Reversed	Reversed
No refusals	87.3	88.0	77.5	Reversed	.07

* Statistically significant.
† By reversed is meant that the adjusted normals have the lower percentage.
‡ As defined by Munroe [10].

justed normals" than by maladjusted normals or patients.

Summary

Selected percentile scores were presented for the major scoring categories of Rorschach protocols of 151 white males, age 17-36, all of whom had had at least an eighth grade education. All Rorschach data were obtained and scored by one psychologist, and the testing was done under uniform test conditions and set.

These median scores for the total "normal" group were compared with the 50th percentile ranks obtained by Cass and McReynolds on comparable scoring categories. Our sample was found to have a somewhat higher median number of responses, which was reflected mainly in the F category, but for the most part the obtained scores were very similar, giving encouragement to the idea that stable and meaningful norms can be obtained for the Rorschach. Our sample confirmed their finding that FM is typically greater than M and CF greater than FC in the normal protocol. The latter inequality is also confirmed by Beck's data.

Since most Rorschach interpretations are based on the concept of normal in the sense of well adjusted, rather than average, it was felt that the most meaningful norms to be obtained from the present sample would be those based on screened normals. All subjects had been interviewed by a psychiatrist and rated as to over-all adjustment. The psychiatric rating was used to divide the sample into adjusted and maladjusted normals, and the 20th 50th, and 80th percentile scores were presented for these subgroups. In addition, the scores of 40 patients with symptoms of manifest anxiety were presented for comparison. While some commonly accepted signs tended to differentiate the maladjusted and/or the patients from the screened normals, many others did not hold for this sample. In particular, the adjusted group used K and m to as great an extent as did the anxious patients. The middle 60% of the group showed a wider range of

$W\%$, $dr\%$, $S\%$, and FM than is usually considered optimal.

Color shock and shading shock as defined by Munroe were found to some extent in all but a very few protocols and thus, as presently defined, are of no value in detecting maladjustment.

It appears that a great deal more empirical research is needed to determine meaningful norms for the Rorschach. In particular, the influence of such factors as sex, race, age, education, and socioeconomic background must be determined, and separate norms obtained when necessary in order to take these factors into account. Also, since Rorschach interpretation is, in general, based on the concept of the optimal or well-adjusted Rorschach protocol, it appears reasonable to obtain norms for screened normal groups. The present paper makes a contribution in this direction by indicating the median score and middle 60% range for a group of young, psychiatrically screened, normal, white males of more than average education. It is hoped that comparable data will be published for other relatively homogeneous groups. Such norms would greatly increase the efficiency and validity of clinical interpretation.

Received November 20, 1953.

References

1. Beck, S. J., Rabin, A. J., Thiesen, W. G., Molish, H., & Thetford, W. N. The normal personality as projected in the Rorschach test. *J. Psychol.*, 1950, **30**, 241-298.
2. Bell, J. E. *Projective techniques.* New York: Longmans, Green, 1948.
3. Brussel, J. A., & Hitch, K. S. *An introduction to Rorschach psychodiagnostics in military and civilian psychiatry.* Utica, New York: State Hospitals Press, 1947.
4. Cass, W. A., Jr., & McReynolds, P. A contribution to Rorschach norms. *J. consult. Psychol.*, 1951, **15**, 178-184.
5. Davidson, Helen H. A measure of adjustment obtained from the Rorschach protocol. *J. proj. Tech.*, 1950, **14**, 31-38.
6. Hertz, Marguerite R. *Frequency tables for scoring responses to the inkblot test.* (3rd Ed.) Cleveland, Ohio: Western Reserve Univer. Press, 1951.
7. Keehn, J. D. Rorschach validation. II. The validity of colour shock in the diagnosis of neuroticism. *J. ment. Sci.*, 1953, **99**, 224-234.
8. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1946.
9. Linn, L. The Rorschach test in the evaluation of military personnel. *Rorschach Res. Exch.*, 1946, **10**, 20-27.
10. Munroe, Ruth. The inspection technique: a method of rapid evaluation of the Rorschach protocol. *Rorschach Res. Exch.*, 1944, **8**, 46-70.
11. Ulett, G. A., Gleser, Goldine, Lawler, Ann, & Winokur, G. An experimental investigation of development of an EEG index of anxiety tolerance by means of photic stimulation — its validation by psychological and psychiatric criteria. *USAF Sch. Aviat. Med. Proj. Rep.*, 1952, Proj. No. 21-37-002, Rep. No. 4.
12. Wischner, G. J., Rotter, J. B., & Gillman, R. D. Projective techniques. In S. W. Bijou (Ed.), *The psychological program in AAF convalescent hospitals.* Washington, D.C.: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 15.) Pp. 144-157.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

An Evaluation of the Memory-for-Designs Test[1]

Alvin R. Howard
VA Hospital, Chillicothe, Ohio[2]
and Donald J. Shoemaker
Ohio State University

Attempts at psychometric isolation of patients with the diagnosis "organic brain damage" have been noticeably inconclusive. One of the techniques offered, but to date little reported on, is Graham and Kendall's Memory-for-Designs Test. [1, 2].

Forty-two male, neuropsychiatric patients, evenly divided between experimentals and controls, comprised the sample studied. The experimental group was composed of patients diagnosed both as "psychotic" and "organic"; the control group consisted of persons diagnosed as "psychotic" only. In selecting subjects, the criteria of the test authors were employed. Individual matching was accomplished wherever possible. Compared with the male and female *S*s employed by Graham and Kendall, the present sample is older, completed fewer grades of formal schooling, and does not contain as wide a variety of diagnoses.

Independent scoring of individual designs (on training protocols) by the authors resulted in a 93% agreement, lending favorable support to Graham and Kendall's correlation of .99 between independent ratings of total scores.

Mean visual-motor raw and difference scores differentiated experimental and control groups beyond the 5% level, also supporting earlier findings of the test authors, although to a lesser degree. Similarly supportive were various correlations duplicated by the present investigators. The Pearson *r*'s between visual-motor score and age, education, and vocabulary, respectively, were .26, —.31, and —.31. The multiple *r* between visual-motor score and age, vocabulary, and grade was .44; between visual-motor score and age and vocabulary was .42. The corresponding coefficients reported by Graham and Kendall were comparably low. Less supportive results occurred in relation to the distribution of visual-motor scores. In respect to both raw and difference scores, a much larger percentage of our control group subjects fell above the critical scores denoting brain damage. Whereas only 4% of Graham and Kendall's "normative adults" had difference scores assigning them to the brain-damage category, our corresponding figure was 19%; similar low percentages were noted by them on other control groups when raw scores were used, whereas 28% of our controls were selected by this method. Experimental group findings by both raw and difference methods are more nearly alike, our respective 52% and 43% approximating their 45%, 50% and 50%. Our results thus suggest the use of increased caution in attributing brain damage to individual neuropsychiatric patients scoring high by either the raw or difference method.

Brief Report
Received March 23, 1954.

References

1. Graham, Frances K., & Kendall, Barbara S. Performance of brain-damaged cases on a memory-for-designs test. *J. abnorm. soc. Psychol.*, 1946, **41**, 303-314.
2. Kendall, Barbara S., & Graham, Frances K. Further standardization of the Memory-for-Designs Test on children and adults. *J. consult. Psychol.*, 1948, **12**, 349-354.

---

[1] An extended report of this study may be obtained without charge from Alvin R. Howard, Veterans Administration Hospital, Chillicothe, Ohio, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4254 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[2] From the Veterans Administration Hospital Chillicothe, Ohio.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

Determinant Shift of Psychoneurotic and Psychotic Patients[1]

Robert G. Gibby and Bernard A. Stotsky
VA Regional Office, Detroit, Michigan

In a previous paper [2], three areas of the relationship of the Rorschach free association and inquiry were investigated. These were: (*a*) their similarity; (*b*) the comparative effect of the stimulus value of the examiner on each; and (*c*) the relationship of the personality characteristics of the subject to changes from free association to inquiry.

Briefly summarized the findings were:

1. The free association record differs significantly from that of the inquiry.
2. The free association is less susceptible to examiner influence than the inquiry.
3. Personality characteristics, as measured by the free association of the Rorschach, of those individuals changing the most differ significantly from those changing the least.

These findings suggested that there were dimensions of the Rorschach worthy of investigation other than those presently tapped by the traditional scoring method. The changes in Rorschach determinants from free association to inquiry, which we shall call "Determinant Shift," especially offered promise. These changes might very well be related to the capacity of a subject to respond to new stimuli and to interact in a structured interpersonal relationship. For instance, the degree of shift of determinants has been used successfully to predict the ability of a subject to form a lasting therapeutic relationship [3].

The calculation of the Determinant Shift score is a relatively simple matter. First, the free association responses are scored independently of the material elicited on the inquiry. Next, the combined free association and inquiry responses are scored. These scorings produce two psychograms, one for free association and one for the inquiry. The score for each determinant on the free association is then subtracted from the score following inquiry. This results in a shift score for each determinant. The Determinant Shift score for all determinants except *F* will usually be either zero or positive, whereas for *F* the score will almost always be zero or negative since the shift is most often from *F* to some other determinant. Consequently, the shift in *F* may usually be regarded as reflecting the total amount of shift.

The administration of the Rorschach is made up of two interdependent, yet quite different, procedures. The first, the free association, is relatively unstructured. The subject is simply requested to free associate to the Rorschach cards as they are presented to him. The second, the inquiry, is more highly structured. Here the subject is asked to explain and to clarify the previously given associations, to delineate the area of the blot utilized and to describe what characteristics of the blot suggested the response. The inquiry itself requires greater attention to, and much more direct interaction with, the examiner. It may be hypothesized that the more sensitive the individual, the more responsive he will be to the examiner. We would then expect greater shift to be shown on determinants from free association to inquiry by more sensitive and more reactive subjects.

To test this hypothesis, a group of 40 psychoneurotic outpatients at a Veterans Administration clinic was compared with a group of 40 psychotic inpatients at a Veterans Administration general, medical, and surgical hospital. The specific prediction was that the psychoneurotics would show greater shift on Rorschach determinants from free association to inquiry than would the psychotics. It was also predicted that they would show greater shift on variables supposedly more indicative of anxiety (*FY*, *Total Y*) than on those presumably related to the more stable aspects of the personality, such as color and movement.

---

[1] From the VA Regional Office, Detroit, Michigan.

The Rorschach determinants investigated were *F*, *M*, *FC*, *CF*, *C*, *FY*, *YF*, *Y*, *V*, *Total Y*, and *Total C*. Only those determinants were studied which yielded large enough frequencies to warrant analysis. Beck's scoring techniques were employed for scoring these determinants. Indices of determinant shift were then calculated for each record [1]. Reliability of the scoring was determined. For this purpose, 10 records previously scored by one of the authors were randomly selected from the entire sample and then rescored by a second person who had been trained previously in the scoring techniques used in the study. Pearson product-moment correlations between the two scores were computed for each determinant, both for the free association by itself and for the total response. For free association the mean reliability was .90; for total response the mean reliability was .94.

In order to control the variability possibly introduced by *R*, an attempt was made to match each psychotic record *exactly* with a psychoneurotic record for the total number of responses. Out of 240 available psychoneurotic records, perfect matches were possible for 37 of the 40 psychotics. For two of the remaining three, there was a difference in *R* of one response. For the other, the difference was three responses. The total difference of *R* for both groups was three. Each record was compared with its matched record for differences in determinant shift. Following this, two types of statistical analyses were performed. The nonparametric "sign" test was used. Here, for the 40 matched pairs of records, the number of times the shift in determinants from free association to inquiry in one group exceeded that of the other was compared for significance. Where the assumption of normality for the amount of shift was tenable, the *t* for correlated means was also utilized to compare differences on shift for each of the determinants.

In order to check the possibility that the shift from free association to inquiry might be a function of existing differences on free association, the two groups were compared on all determinants in the free association. No significant differences were found for any of the determinants.

Results

A previous study [2] demonstrated that the determinant shift for *F* may serve as an indicator of the shift in all determinants from free association to inquiry. This index was used in the present study to compare the total shift of the psychoneurotic with that of the psychotic group. The psychoneurotic groups showed, using the sign test,[2] more change, with the difference in the index of determinant shift for *F* being significant at the .001 level. The prediction that the psychoneurotic group would change more from free association to inquiry was thus upheld.

The indices of determinant shift for the shading variables (*FY* and *Total Y*) in each group were also compared by the sign test. The differences between the groups were significant for both variables. For *FY*, the difference was significant at the .01 level; for *Total Y*, at the .001 level.

The differences between the indices of determinant shift for the color and movement variables were not found to be statistically significant. The index of shift for color tended to be higher, and that for movement lower, in the psychoneurotic group. The results of the comparison of the groups by determinants are summarized in Table 1.

---

[2] The first test of the hypothesis was by means of the sign test. A problem presented itself in terms of what to do with cases where the amount of shift from free association to inquiry was the same for both groups. The following procedure was finally adopted: The results of the sign test were plotted on a normal curve; first, going from left to right, the proportion of cases where the shift was greater for psychotics; then the ties; and finally the proportion of cases where the shift was greater for neurotics. Next, the distance between the *z* scores for the ties was found. Then, with the mean set at zero, half of the distribution of ties was plotted on one side and half on the other side of the mean. From that point, determining the theoretical frequencies of the ties was simply a matter of consulting the table of normal curve functions and adding the area from *m* to *z* on both sides of the mean. The proportion of the curve thus obtained was multiplied by 40 for the theoretical frequency of ties. The remaining area on each side of the curve was multiplied by 40 to give the theoretical frequency for the cases where psychotics showed more shift than neurotics and for cases where neurotics showed more shift than psychotics. The chi-square values (two degrees of freedom) between obtained and theoretical frequencies were then computed for the three signs.

Table 1
Comparison of Neurotics and Psychotics on "Determinant Shift" by the "Sign" Test

Determinant	P > N*	P = N**	N > P†	*p*
F	8	5	27	.001
M	12	21	7	.46
FY	7	8	25	.01
YF	3	31	6	.64
Y	—	39	1	—
V	3	31	6	.64
Total Y	5	9	26	.001
FC	7	16	17	.12
CF	7	21	12	.46
C	—	—	—	—
Total C	9	11	20	.11

* Instances in which psychotics shifted more than neurotics from free association to inquiry.

** Instances in which the amount of shift for neurotics and psychotics was equal from free association to inquiry.

† Instances in which neurotics shifted more than psychotics from free association to inquiry.

The mean differences in determinant shift for *F*, *FY*, *Total Y*, and *Total C* were then compared by means of the *t* test. The *t* test was not used for the other determinants inasmuch as the distributions were markedly skewed by the presence of many 0 scores. These results are summarized in Table 2.

Table 2
Comparison of Neurotics and Psychotics on "Determinant Shift" by the *t* Test

Determinant	Means: Psychotics	Means: Neurotics	Diff.	*t*	*p*
F	−3.15	−5.15	2.00	2.78	.01
FY	1.23	2.38	1.15	3.19	.01
Total Y	1.42	2.93	1.51	3.37	.001
Total C	1.38	1.85	.47	1.15	.25

The findings for *t* agree very closely with those for the "sign" test. Neurotics show significantly greater shift for *F*, *FY*, and *Total Y*. They also show more shift for the color variables, but again, as for the sign test, not significantly.

From these findings it can readily be seen that the Determinant Shift scores for *F*, *FY*, and *Total Y* differentiate at a significant level of confidence between neurotics and psychotics in the predicted direction. Fifty-six out of 80, 70 per cent of the patients, are correctly identified when the median for the shift in *Total Y* for both groups is used as a point of division, with scores above the median classified as neurotics and scores below the median as psychotics. The number of cases correctly classified is significantly better than chance. Their greater ability to shift suggests that psychoneurotics show more sensitivity to the testing situation and respond to the additional stimuli of the inquiry more actively than do psychotics. The greater increase in shading from free association to inquiry among psychoneurotics may be indicative of the presence of more anxiety in their reactions than in those of the psychotics.

Among the determinants of the Rorschach, the amount of shift may well be the important variable differentiating psychotic and psychoneurotic patients. The present-day method of scoring the total response, including both free association and inquiry, tends to obscure such shifts. Examination of our data demonstrates the point clearly. Comparison of the two groups by the traditional method of scoring determinants shows significant differences only on the shading variables *FY* and *Total Y*. On the other variables the shift from free association to inquiry leads to a convergence of the groups. For example, neurotics show more *F* on free association than psychotics. As a result of the shift from free association to inquiry, psychotics end up with more *F*. Similarly, psychotics show slightly more *FC*, *CF*, *C*, *V*, *YF*, *Y*, and *M* on the free association; as a result of the shift, psychoneurotics end up with slightly higher scores for these variables. The differences which exist between the two groups in the amount of shift, however, tend to be concealed; in the case of *F*, a very significant difference.

Comparison of psychotics with most and least shift. In the previous study, in comparing the 40 psychoneurotics who shifted most with the 40 who shifted least, significant differences were found for *F*, *M*, and *FC* on free association. The tentative conclusion drawn from the data was that the group shifting least was more comfortable in the interpersonal situation and less ready to interpret the situation as one of threat. The question arose as to whether the same held true for psychotics. Would a group of psychotics who shifted little compared with a group who shifted considerably show the same differences on *M*, *F*, and *FC*?

To test this hypothesis, 16 psychotics giving shift scores of 0 and 1 were compared with 15 giving shift scores of 5, 6, and 7. Since there was a significant difference for *R* between the two groups, the arc-sine transformation was used to control for *R*. Fisher's exact test was then applied to the data, the fourfold categorization being in terms of presence or absence of the determinant on free association. Significant differences at the five per cent level of confidence were found only for *M*, *CF*, and *C*, with the group that showed the most shift giving more *M* and *CF* responses, and the group that showed the least shift giving more pure *C* responses.

The obvious discrepancy between the findings for psychoneurotics and those for psychotics does not appear to be due to chance in view of the significance of the differences found for the two groups. It may be that the meaning of shift for psychotics is different than for psychoneurotics. Among psychotics, shift may be indicative of retention of better emotional and intellectual control. Patients who do not shift may have become unable to respond effectively to the situation. Among neurotics, on the other hand, shift may be indicative of some loss of control, while lack of shift may be characteristic of a person with brittle, rigid, yet intact, defenses. The use of external criteria of adjustment and of retention of control would help shed light on the meaning of shift for the two groups. At any rate, in the absence of validating evidence, any conclusions reached here must be tentative. Work is in progress to validate the findings of the present study on a larger group of psychotics selected from patients at a neuropsychiatric hospital.

Summary

In this study a new dimension of the Rorschach, the determinant shift from free association to inquiry, was applied to a group of psychotics and psychoneurotics.

The specific predictions were:

1. Neurotics would show greater shift from free association to inquiry than psychotics.

2. Neurotics would show greater shift on variables presumably indicative of anxiety, *FY* and *Total Y*. Using both the *t* test and the sign test, both predictions were confirmed at very significant levels of confidence.

It was further demonstrated that the Determinant Shift score emphasized differences between the two groups that the classical method of scoring the determinants tended to obscure.

Sixteen psychotic patients who shifted least were compared on the free association scores with 15 psychotic patients who shifted most. The group which showed most shift gave significantly more *M* and *CF*, while the group which showed least shift gave significantly more pure *C*. The findings for psychotics were compared with those for psychoneurotics, and the implications of these findings were discussed.

In view of the significant findings of this and other studies, the Determinant Shift score appears to be suitable for further experimental investigation and validation.

Received August 31, 1953.

References

1. Gibby, R. G. Examiner influence on the Rorschach protocol. *J. consult. Psychol.*, 1952, **16**, 449-455.
2. Gibby, R. G., & Stotsky, B. A. The relation of Rorschach free association to inquiry. *J. consult Psychol.*, 1953, **17**, 359-364.
3. Salk, L. The relationship of elaboration on the Rorschach inquiry to continuance in psychotherapy, Unpublished doctor's dissertation, Univer. of Michigan, 1953.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

Rorschach Card Rejection in Psychodiagnosis

Ivan N. Mensh and Joseph D. Matarazzo
Washington University School of Medicine

The phenomenon of rejection of the stimulus cards of the Rorschach test has been observed since Rorschach's early work with this diagnostic method. He noted the differential significance of card rejection by normal, neurotic, "feebleminded hysterical," and schizophrenic subjects:

> *Failure to answer.* Normal subjects almost never fail to give a response for the several plates. Occasionally neurotic subjects fail to answer; this is caused by inhibitions due to complexes. Feebleminded hysterical subjects often reject the figures because they are afraid they will give stupid answers; we deal with an "intelligence complex." Failure frequently occurs in a typical manner in schizophrenics, even in latent or practically recovered cases; they suddenly refuse to answer, even though they have been giving many and good responses all along. The refusal may come with figures not ordinarily considered difficult. The blocking may be insurmountable although in all other kinds of cases an answer may be obtained on coaxing [8, p. 22].

Beck, in his early monograph [1], also summarized the psychological meaning of card rejections by the schizophrenic, depressed, neurotic, and feebleminded, relating the behavior to "blocking, negativism, or stuporous condition" in the first group; to "generalized inhibition and self-depreciation" in the second; to "self-distrust and lack of confidence" in neurotics; and to "criticism" in the feebleminded. Klopfer and Kelley [6] present greater detail and occasional illustrations of card rejection; Bell [3] examines the "main indications of rejections of the various plates"; Yacorzynski [9], Zubin [10], and Zubin and Young [11], among others, make passing reference to the phenomenon of Rorschach card rejection by subjects.

Review of these reference texts in the Rorschach method does not indicate, however, any statistical basis for differential significance of card rejection by the various diagnostic groups seen in clinical practice. Further, a study of the indexes of *Psychological Abstracts* since 1946 and of the *Rorschach Research Exchange and Journal of Projective Techniques* since its first volume yielded but a single article [5] devoted solely to card rejection in Rorschach testing, and this was in a Swiss journal. Although a number of studies have reported card rejection as part of the findings, none has been designed specifically to investigate systematically the frequency of rejection of the stimulus cards. The present report is presented as an attempt to offer data on Rorschach card rejection and its relationship to other personality characteristics.

Research Design

Because of the significance and use of card rejection in interpretation of data gathered by the Rorschach method, the present study was undertaken to investigate the statistical incidence of this phenomenon and to seek for correlative data. The sample of subjects studied was a total sample during a two-year period, and consisted of 201 patients given psychological examinations in the neuropsychiatric service of an inpatient hospital. Of the total sample, 100 were psychoneurotic (*PN*), 74 were psychotic (*PS*), and 27 were brain-damaged (*OR*), as diagnosed, independently of the Rorschach data, by psychiatric and psychological staff members. Data were collected on the incidence of rejection of stimulus cards of the Rorschach, and on other characteristics of the sample studied, including age, sex, Wechsler IQ, education, *R*, $T/1R$, and T/R. An analysis of the reaction time characteristics and their significance has appeared elsewhere [7].

Results

Identifying data and Rorschach response characteristics of the subjects studied are summarized in Table 1. The differences

Table 1

Rorschach and Other Identifying Variables among the Three Samples Studied

Variable	Psychoneurotic (*PN*)		Psychotic (*PS*)		Organic (*OR*)	
	Respond	Reject†	Respond	Reject	Respond	Reject
N	76	24	60	14	17	10
Age						
Mean	30.62	30.64	27.28	28.78	37.53	40.89
SD	9.60	9.24	11.25	8.78	15.27	11.14
Sex						
Male	32	12	26	7	10	8
Female	44	12	34	7	7	2
Education						
Mean	13.40**	10.33	12.92	11.08	11.88	8.57
SD	2.79	3.78	3.18	4.01	4.30	1.18
IQ (W-B)						
Mean	113.47**	99.90	104.18	96.00	94.33	90.50
SD	12.07*	19.63	17.57	25.17	20.35*	8.17
R						
Mean	39.20**	16.00	55.47**	19.78	29.88*	13.00
SD	21.14**	7.96	50.94*	7.60	27.09**	4.37
T/R						
Mean	41.54**	65.33	42.17	72.15	72.76	61.00
SD	27.86	31.92	24.42**	65.12	49.64	26.42
T/1R						
Mean	18.00	25.54	16.68*	31.36	28.24	22.89
SD	14.62	19.75	15.10*	25.11	22.78	18.61

* Responding and rejecting samples significantly different at the .05 level.

** Responding and rejecting samples significantly different at the .01 level.

† Patients rejecting one or more cards.

among the three samples with respect to Rorschach response characteristics and the other identifying data were reported earlier [7]. In brief, the mean age of the organic sample was significantly greater and their mean Wechsler-Bellevue IQ was significantly less than the means of the two other groups. These differences may be the result, at least in part, of lowered functioning due to the cerebral pathology of this group. The neurotic and psychotic groups also had more years of schooling. Further, among the patients diagnosed as suffering from organic pathology there were significantly more men than women (67 per cent males, contrasted with 44 to 45 per cent in the other two samples). Thus, the three samples were not matched groups in *N*, age, sex, education, or intelligence level; this raises the problem of significance of variation among the Rorschach characteristics of the samples. Statistically significant differences among the mean values of Rorschach responses may possibly be attributed to one or more of the identifying variables listed above, and not to a disease category. Therefore, a series of correlations was necessary to establish the extent of relationship among the identifying variables and the response variables. These were computed and are reported in Table 2.

It may be seen that in none of the samples did age correlate significantly with other factors, indicating that this variable probably was not significant in the varying Rorschach response characteristics, though the *OR* group was significantly older than groups *PN* and *PS*. However, both education and intelligence level did correlate significantly with certain of the other response characteristics, but there was not a consistent relationship from sample to sample except between education and IQ, and here the correlations were of the same order: .63 to .66. There was a positive correlation between education and total number of responses (*R*) for the *PN* and *OR* groups but not for the *PS* group; yet only in the latter group was there a significant though inverse relationship ($r = —.35$) between education and mean time per first response (*T/1R*).

Table 2

Correlations between Variables among the Three Samples

Variable	Sample *PN* ($N = 107$)	 *OR* ($N = 29$)	 *PS* ($N = 74$)
Age – Educ.	—.07	—.16	—.03
Age – IQ	.22	.21	.08
Age – R	.00	—.12	—.07
Age – T/R	—.03	.30	.20
Age – $T/1R$	.06	.02	.21
Age – ΣT	—.02	.25	.06
Educ. – IQ	.63**	.63**	.66**
Educ. – R	.28**	.49*	.14
Educ. – T/R	—.19	—.15	.00
Educ. – $T/1R$	—.16	—.18	—.35**
Educ. – ΣT	.09	.37	.09
IQ – R	.32**	.18	.25
IQ – T/R	—.29*	—.07	—.11
IQ – $T/1R$	—.23	—.28	—.19
IQ – ΣT	.07	.19	.02
R – T/R	—.41**	—.30	—.22
R – $T/1R$	—.32**	—.11	—.17
R – ΣT	.54**	.74**	.75**
T/R – $T/1R$	.60**	.60**	.70**
T/R – ΣT	.34**	.34	.25*
$T/1R$ – ΣT	.30**	.35	.13

* Significant at 5% level of confidence.
** Significant at 1% level of confidence.

Similarly, there was no consistency in correlations between IQ and factors other than education, and there were significant correlations only for IQ – R and IQ – T/R in the *PN* sample. In summary, of the 63 correlations 19 were significant, 16 more at the $p = .05$ level of significance than would be expected from chance variation alone (see Table 2). However, the relationships did not appear consistently, although the relative significance of R is apparent in the 7 of the 19 significant correlations in which R was one of the variables, and in the 8 correlations in which time was a factor. Further, 10 of the 19 significant correlations occurred in the *PN* group, with the remaining 9 about evenly distributed between the other two groups, suggesting the greater relative order of relationship among these variables in neurotics as compared with the more seriously disorganized individuals of the psychotic and organic pathology samples.

Table 3

Rorschach Rejections by Sample, Sex, and Card

Sample and sex	N	Rejections by cards I	 II	 III	 IV	 V	 VI	 VII	 VIII	 IX	 X	 Total
Psychoneurotic												
Male	44	1	3	0	4	2	5	4	1	5	3	28
Female	56	1	2	0	3	2	4	3	1	8	5	29
Total	100	2	5	0	7	4	9	7	2	13	8	57
Psychotic												
Male	33	0	1	0	2	0	3	1	1	4	2	14
Female	41	0	0	0	2	0	2	3	0	0	2	9
Total	74	0	1	0	4	0	5	4	1	4	4	23
Organic												
Male	18	0	0	1	1	1	2	3	0	4	2	14
Female	9	0	0	0	2	0	0	1	0	1	1	5
Total	27	0	0	1	3	1	2	4	0	5	3	19
All groups												
Male	95	1	4	1	7	3	10	8	2	13	7	56
Female	106	1	2	0	7	2	6	7	1	9	8	43
Total	201	2	6	1	14	5	16	15	3	22	15	99

In Tables 3 and 4 are presented the data on Rorschach card rejections in the total sample, tabulated by disease category, sex, card, and individual. Although twice as many organics, proportionally, as psychotics rejected cards (37 per cent and 19 per cent, respectively), and as many as 24 per cent of the sample diagnosed as psychoneurotic, the value of p for the chi-square statistic was between .20 and .10, not significant at the usually accepted level of .05 or lower. The mean number of cards rejected was variously 1.64 (*PS*), 2.38 (*PN*), and 1.90 (*OR*). Among the 201 patients in the total sample there were 99 rejec-

Table 4

Rorschach Rejections by Sample, Sex, and Individual

Sample and sex	*N*	Frequency of rejection 1	 2	 3	 4	 5	 6	Total persons
Psychoneurotic								
Male	44	2	5	1	2	1	0	11
Female	56	6	2	3	1	0	1	13
Total	100	8	7	4	3	1	1	24
Psychotic								
Male	33	4	0	2	1			7
Female	41	5	2	0	0			7
Total	74	9	2	2	1			14
Organic								
Male	18	4	0	2	1			7
Female	9	2	0	1	0			3
Total	27	6	0	3	1			10
All groups								
Male	95	10	5	5	4	1	0	25
Female	106	13	4	4	1	0	1	23
Total	201	23	9	9	5	1	1	48

tions of stimulus cards by 48 individuals, a mean rate of 2.06 for those who did reject. These rejections were principally of cards IV, VI, VII, IX, and X. Similarly, Klopfer and Kelley [6] reported rejections most often of cards II, IV, VI, and IX. In the present study 24 per cent of the total sample rejected cards, 50 per cent of the stimulus cards provoked 83 per cent of the rejections, and in these rejections 23 of the 48 patients rejected only one card and only two rejected more than four cards. Brockway *et al.* [4] reported rejection by 22.5 per cent of their neurotic patient sample, an incidence consistent with our findings. Hiltmann [5] also concluded from a review of Rorschach literature that card rejections "are rather frequent."

Tests of significance of differences between means and their sigmas for responding and rejecting patients are summarized in Table 1, which also reported the sample characteristics. The "reject" columns represent those patients who rejected one or more stimulus cards. There were no significant differences between the subsamples, within each diagnostic category, with respect to age and sex. On the education variable, only the *PN* group showed higher education for the responding sample than for those rejecting Rorschach stimulus cards. Since education and Wechsler-Bellevue IQ were correlated significantly (*r*'s of .63 to .66) for each of the three diagnostic samples, only among the *PN* group did the responding patients have higher IQ's, as would be expected. However, variability differed significantly among both the *PN* and *OR* samples, but in reverse directions — greater variation among those *PN* patients rejecting cards, but less variation among rejecting *OR* patients. As would be expected from the consequences of card rejection, for all samples there were significantly fewer responses and correspondingly decreased variation among the patients rejecting cards. Yet, only in the *PN* sample was there a significant mean difference for *T/R*. The patients of the *PS* sample who rejected cards were far more heterogeneous with respect to the *T/R* factor, so that the apparent mean difference of 30 seconds was not significant. But in this sample alone were there significant differences between means and sigmas in the analysis of data on the *T/1R* variable.

Summary

Rorschach card rejection has been observed and reported since the introduction of the Rorschach technique for psychodiagnosis. Differential significance of the phenomenon of card rejection frequently has been suggested as a diagnostic aid, but a systematic study of the statistical incidence of this behavior has not been reported. In the present study, 201 hospitalized patients were examined with the Rorschach method. In the total sample there were 100 psychoneurotic, 74 psychotic, and 27 brain-damaged patients. The three groups were not matched but represent a typical sample of patients seen in clinical psychologic practice on an inpatient service. Intercorrelations among variables representing sample characteristics were not consistent, but suggest that the relative order of relationships among these characteristics in neurotics is higher than in the samples of the more seriously disorganized patients diagnosed as psychotic or brain-damaged.

Rejection of one or more of the Rorschach stimulus cards was found in 19 to 37 per cent of the samples; five of the ten cards provoked 83 per cent of the rejections; in the responses of subjects who rejected the cards, on the average, two cards were rejected per patient, although about half of the rejecting individuals rejected but one card and in only two cases were more than four cards rejected. A number of trends that differentiated patients responding to all of the stimulus cards from those rejecting one or more of the cards appeared, but there were few statistically significant differences. As in an earlier study, the major finding with respect to Rorschach card rejection is the extreme variability in response characteristics, emphasizing again the care necessary in utilizing data gathered by the Rorschach method to interpret behavior.

Received October 20, 1953.

References

1. Beck, S. J. *Introduction to the Rorschach method.* New York: Amer. Orthopsychiat. Ass., 1937.
2. Beck, S. J. *Rorschach's test.* Vol. I. *Basic processes.* New York: Grune & Stratton, 1944.
3. Bell, John E. *Projective techniques.* New York: Longmans, Green, 1948.

4. Brockway, Ann L., Gleser, Goldine C., & Ulett, G. A. Rorschach concepts of normality. *J. consult. Psychol.*, 1954, **18**, 259-265.

5. Hiltmann, Hildegard. Das Versagen im Rorschachtest. (Rejections in Rorschach testing.) *Beih. Schweiz. Z. Psychol. Anwend.*, 1948, No. 13, 123-130. (*Psychol. Abstr.*, 1948, **22**, No. 3941.)

6. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1946.

7. Matarazzo, J. D., & Mensh, I. N. Reaction time characteristics of the Rorschach test. *J. consult. Psychol.*, 1952, **16**, 132–139.

8. Rorschach, H. *Psychodiagnostics.* (3rd Ed.) New York: Grune & Stratton, 1942.

9. Yacorzynski, G. K. *Medical psychology.* New York: Ronald, 1951.

10. Zubin, J. *Quantitative techniques and methods in abnormal psychology.* New York: Columbia Univer. Bookstore, 1950.

11. Zubin, J., & Young, K. M. *Manual of projective and cognate techniques.* Madison: College Typing Co., 1948.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

Intelligence Test Reliability for Cerebral Palsied Children[1]

David H. Crowell
University of Hawaii
and Doris C. Crowell
Honolulu, T. H.

Cerebral palsied children having at least two psychological examinations using the same instrument for the initial and final contact, and having mental ages definitely stated, were selected from the register of the Board of Health, Honolulu, Hawaii, and the Psychological Clinic, University of Hawaii. Individual diagnoses of the 61 cases selected included the full range of motor disabilities as well as muscular involvements. Examinations, ranging from 2 to 6 per individual, were based upon the Stanford-Binet or Porteus Composite Binet [1], 49 cases; the Cattell Infant Intelligence Scale, 11 cases; and the Merrill-Palmer Scale of Mental Tests, 1 case. Seventeen different psychologists were involved in 46 cases and the same examiner in 15 cases.

Based upon the median of all test-retest intervals, two groups were formed: Group I, 16 males and 12 females with an average test-retest interval of 15 months, and Group II, 20 males and 13 females with an average test-retest interval of 43 months. The total group had an average test-retest interval of 30 months.

This sample was highly skewed, with roughly 50 per cent of the group classifiable as severely retarded. Analysis of the total retest results showed that approximately 51 per cent of the total group deviated within ± 5 IQ points of the initial test score and approximately 75 per cent within ± 10 IQ points of the initial test score. Pearsonian coefficients of reliability, based upon mental ages, and confidence limits at the 1 per cent level are as follows: Group I, $r = + .98$ with confidence limits of $+.94$ and $+.99$; Group II, $r = +.78$ with confidence limits of $+ .52$ and $+ .91$; and total group, $r = + .92$ with confidence limits of $+ .87$ and $+ .95$. Standard errors of estimate are, respectively, ± 7, ± 13, and ± 11.

The data suggest that the mental age obtained on the initial examination of a cerebral palsied child can be accepted as highly reliable for ordinary diagnostic purposes, provided the clinical psychologist has adequate training and experience. If errors of clinical judgment on the initial examination occurred, they erred on the side of caution. Moreover, these coefficients of stability are highly significant and have wide generality since there were increased chances for the operation of errors inherent in the measuring instrument, errors due to clinical judgment, and errors resulting from fluctuations in ability because of maturation.

Brief Report.
Received March 1, 1954.

Reference

1. Jorgensen, Margaret. A comparative study of the Stanford-Binet Form L and the Porteus Composite Binet. Unpublished master's thesis, Univer. of Hawaii, 1951.

---

[1] An extended report of this study may be obtained without charge from David H. Crowell, Department of Psychology, University of Hawaii, Honolulu 14, Hawaii, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4287 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for 35 mm. microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

A Cross Validation of Davidson's Rorschach Adjustment Scale

Raymond J. Corsini
University of Chicago

and Harold F. Uehling
Wisconsin Department of Public Welfare

Among the objective scales based on the Rorschach protocol are those constructed by Buhler [1], Davidson [2, 3], Munroe [9], Hughes [6], and Piotrowski [10]. The first three are of particular interest since they attempt to measure general adjustment. These scales were constructed along the general principles employed by objective test constructors: two populations believed to be different in terms of a criterion were established; various signs were tested to determine their discriminative powers; signs that met certain limits of significance of difference were isolated, and were then combined to form the scale. However, the crucial test of the value of any scale is its ability to discriminate between new populations.

The purpose of this study is to test the Davidson scale for generality of measurement. The Davidson scale was selected for several reasons: first, the signs are limited to 17, 15 of which are so objective that a clerk might score them; second, previous research has been done with this scale to indicate that it is reliable [2, 3]; third, the scale has been demonstrated capable of differentiating between "adjusted" and "maladjusted" school children [2, 5, 8].

The objectification of the Rorschach by lists of signs has two purposes: (*a*) validation of hypotheses relating to various signs, and (*b*) the possible scaling of various degrees of emotional maladjustment. Most Rorschach workers are generally dubious about the value of such scales, believing that projective instruments reveal unique and ineffable aspects of personality and that the quantitative conversion of qualitative aspects of the personality yield little of extra value. However, it is possible that objective procedures have some value. For example, in comparing two clear-cut populations, one "normal" and one "abnormal," if it is found that differences with regard to any sign are nonsignificant or even in a direction not predicted, then the clinical insights that suggested the interpretation of these signs may be questioned.

Method

The two populations are called "Guard" and "Inmate." Both consist of males, matched for age. Since it was not possible to match for intelligence, only protocols of inmates who were known to have at least average intelligence were selected; also the guards had the equivalent of high school graduate standing, and were presumably at least average in intelligence. The guard population is composed of men accepted as probationary prison guards at the Wisconsin State Prison at Waupun. They ranged in age from 21 to 45, and were of undamaged reputation. Each probationary guard had passed, in addition to the usual civil service requirements, the criterion of an interview by the appointing authorities. Therefore, they form a screened population since people who would appear unsuitable in terms of personality defects or handicaps of adjustment would not be hired. No evidence exists to support a belief that this group as a whole differs from any normal sample of the adult male population.

The inmate population consists of men committed to prison for felony crimes. This alone might be considered to be presumptive evidence that the group as a whole is abnormal, but the population actually subjected to the Rorschach was a selected sample of this larger group. Only those inmates who, in the opinion of a psychiatrist, appeared seriously emotionally conflicted were referred to the psychologist for projective analysis. Therefore, the sample of inmates who were matched for age against the guards was doubly distilled for abnormality: first by the court, and second by the initial psychological and psychiatric examinations.

All Rorschachs in this study were administered by Uehling. Analysis was done by both writers, and for the two signs that called for judgment, *no color shock* and *no shading shock* (signs 15 and 16), independent evaluations were made. If agreement did not occur, the protocols were examined more carefully until an agreement was reached. The analysis of the protocols was done without knowledge of the class identity of the subject. The frequency of occurrence of the Davidson signs was tabulated for each population. The significance of the difference was obtained by dividing the percentage difference by the standard error of the percentage of the difference from Edwards' formula 21 [4, p. 77]. The resulting t's were converted into fiducial limits of probability using the two-tailed test on the assumption that none of these signs is significant in either direction [7]. The mean and standard deviation of the scale for each of the populations were also calculated.

Table 1
Occurrence of Favorable Signs for Two Populations of Prison Inmates and Guards
($N = 50$ for each population)

Sign	Inmates	Guards	t	p
1. $M > FM$ or $M = FM$	26	27	.02	.9
2. M, 3 or more	8	13	1.23	.3
3. *Sum* $C > Fc + c + C'$	30	21	1.80	.1
4. $F\%$ — 50 to 20	21	18	.75	.5
5. $Dd + S$ 10% or less	32	34	.04	.9
6. P, 4 or more, up to 30% of R	33	38	1.11	.3
7. R more than 20	25	30	1.01	.4
8. $FC > CF$ or $FC = CF$	25	26	.02	.9
9. FC 2 or more	21	30	1.80	.1
10. No pure C	23	31	1.61	.2
11. %8, 9, 10-40 to 60	17	14	.65	.6
12. $FK + Fc$, 2 or more	12	11	.02	.9
13. $W{:}M = 2{:}1$	12	9	.75	.5
14. $A\%$ 50 or less	30	37	1.49	.2
15. No color shock	23	31	1.61	.2
16. No shading shock	37	34	.75	.4
17. No refusals	35	45	2.50	.02

Results

Table 1 lists the 17 signs, the frequency of their occurrence in both groups, the t tests, and the levels of significance. It will be noted that 14 of the 17 signs are in the expected direction. However, the "abnormal" population does "better" than the "normal" population on signs 3, 4, and 16. Sign 3, *Sum C greater than Fc plus c plus C′*, is significant at the 10% level. Of the remaining 14 signs, twelve are insignificantly in favor of the "normal" group. Sign 9, *FC two or more,* is significant at the 10% level, and sign 17, *no refusals,* is significant at the 2% level.

The mean Davidson score for the 50 inmates is 8.24 ± 2.32, and for the guards it is 9.08 ± 2.36. The difference of 0.84 is significant at the 10% level with a t of 1.77. These figures compare favorably with the scores reported by Davidson for various populations [3]. The only adult norm is for seven college students, given as 8.14; this score is nonsignificantly lower than both our inmate and guard norms.

Discussion

Scales may be restricted to special populations. While a scale for length is general for any object, psychometric scales may be specific. Thus an intelligence test may be suitable for only a particular population, and scales for measuring general adjustment with the Rorschach may be restricted in their applicability. In the case of the Davidson scale as applied to our populations, 14 of the 17 signs are in the expected direction, but of the three that meet the 10% level of significance, one of them is in the wrong direction. Only one sign, number 17, *no refusals,* was significant at beyond the 5% level. Whereas only 10% of the guards had any refusals, 30% of the inmates refused one or more cards.

The scale as a whole measures in the expected direction, but not to any significant degree. The gain resulting from using this scale is minimal. We could expect to select eight fewer guards and eight more inmates out of a population of 100 guards and 100 inmates than by chance if we were to use this scale for selection purposes. This is certainly no impressive saving over chance.

Future Research

In a previous study [12] using the same populations, it was discovered that while the reaction times of the inmates and the guards did not differ significantly from each other, both differed significantly from the reaction times reported by Sanderson [11]. The reason for the difference was stated to be the reluctance of the guards and inmates in contrast to the eagerness of Sanderson's group. However, this probable reluctance should not be germane to the present investigation.

It is intended in the future to check the discriminative capacity of other scoring schemes to separate these populations; and, in addition, to determine the ability of psychologists, unaided, to separate the two groups.

Summary

The Rorschach protocols of two populations equated for sex, color, age, and intelligence were analyzed in terms of the signs on the Davidson Rorschach Adjustment Scale. One population consisted of probationary prison guards who took the Rorschach as part of the routine of selection and are considered to be "normal." The other population, consisting of prison inmates convicted of serious crimes, screened by a psychiatrist, and referred to projective testing in terms of suspected personality difficulties, is considered "abnormal."

Of the 17 Davidson signs, only 3 were significant beyond the 10% level, 1 of them in the "wrong" direction. Only one sign, *no refusals*, was significant beyond the 5% level. The scale as a whole was significant at the 10% level. By setting a cutting score at 9.00, an 8% saving over chance resulted, which is believed to be too small to warrant the use of this method for the purpose of separating adults in terms of personality maladjustment.

Received November 30, 1953.

References

1. Buhler, Charlotte, Buhler, K., & Lefever, D. W. *Rorschach standardization studies. I. Development of the basic Rorschach score with manual of directions.* Los Angeles: Rorschach Standardization Studies, 1948.
2. Davidson, Helen H. *Personality and economic background.* New York: King's Crown Press, 1943.
3. Davidson, Helen H. A measure of adjustment obtained from the Rorschach protocol. *J. proj. Tech.*, 1950, **14**, 31-38.
4. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
5. Gair, M. Rorschach characteristics of a group of very superior seven-year-old children. *Rorschach Res. Exch.*, 1944, **8**, 31-37.
6. Hughes, R. N. Rorschach signs for the diagnosis of organic pathology. *Rorschach Res. Exch.*, 1948, **12**, 165-167.
7. Jones, L. V. Tests of hypotheses: one-sided versus two-sided alternatives. *Psychol. Bull.*, 1951, **49**, 43-46.
8. Margulies, H. Rorschach responses of successful and unsuccessful students. *Arch. Psychol., N. Y.*, 1942, No. 271.
9. Munroe, Ruth L. The inspection technique: a method of rapid evaluation of the Rorschach protocol. *Rorschach Res. Exch.*, 1944, **8**, 46-70.
10. Piotrowski, Z. A. The Rorschach ink-blot method in organic disturbances of the central nervous system. *J. nerv. ment. Dis.*, 1937, **86**, 525-537.
11. Sanderson, H. Norms for "shock" in the Rorschach. *J. consult. Psychol.*, 1951, **15**, 127-129.
12. Uehling, H. F. Rorschach "shock" for two special populations. *J. consult. Psychol.*, 1953, **16**, 224-225.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

Failure to Return for Treatment in Tested and Nontested Clinic Patients[1]

Miriam McCue, Morris Goodman,[2] and Melvin Rosenthal[3]
VA Mental Hygiene Clinic, Lowell, Massachusetts

When new clients fail to return to a clinic for treatment after they have been psychologically tested, it is not uncommon to assume that the tests were "too threatening."

To test the validity of this assumption, the following hypothesis was formulated: Giving psychological tests to patients when they first come to a clinic has the effect of increasing the incidence of failure to return for treatment.

A brief battery of tests was given to all clients who reported to an out-patient clinic during a six-month period. The tests were the Kent EGY, Draw-A-Person, Bender Gestalt, and a sentence completion test. The *S*s who had received therapy since discharge from the service were eliminated, as were Korean and World War I veterans in the final tabulation. The experimental group contained 20 *S*s.

Patients who were not tested at the time of intake made up the control groups. There were 51 *S*s who entered during the same months of the previous year, and 10 who entered during the same months of the year following that selected for "intake testing."

The *S*s in all groups were male, white World War II veterans who lived in the city in which the clinic is located, or in nearby communities. Their ages ranged from 23-43.

Only one (5%) of the tested *S*s did not return for treatment, whereas in the two control groups six and one *S*s, respectively, i.e., 11% of the combined control group, did not return. Thus the hypothesis is not borne out. Fewer of the tested *S*s than of the nontested *S*s failed to return for treatment. The difference is not significant, however, and the direction it takes could have occurred by chance variation in the sampling of patients.

Length of time in treatment may be influenced by many factors, of which testing is only one. If the testing situation was threatening to patients when they first came to the clinic, the *S*s in the experimental group should have shown a tendency to remain in therapy a shorter time than the control *S*s. An incidental finding of this study does not support this second hypothesis. The median number of times the tested *S*s returned for treatment was 10, while the medians were 8 and 7.5 respectively for the other two groups. Again the trend seems to be the reverse of the hypothesis. However, the difference between the control and the experimental group medians was not significant, and therefore time in treatment, in terms of therapeutic interviews, does not differ significantly for the tested and the nontested *S*s.

Apparently the testing at this clinic was not threatening enough to deter the patients from returning for treatment, or to reduce the number of their visits. The effects of different tests, different examiners, and different therapists remain problems for further research.

Brief Report
Received May 3, 1954

---

[1] An extended report of this study may be obtained without charge from Miriam McCue, 20 North Hancock Street, Lexington, Mass., or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4251 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[2] Now at VA Mental Hygiene Clinic, Newark, New Jersey.

[3] Now at VA Hospital, Brockton, Massachusetts.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

An Evaluation of Psychological Reports In a Clinical Setting[1]

Sol L. Garfield
VA Hospital, Downey, Illinois
Ralph W. Heine and Morton Leventhal
VA Mental Hygiene Clinic, Chicago, Illinois

The clinical psychologist currently spends a large share of his time administering psychological tests to patients and writing fairly extensive reports of his findings. These reports are usually presented at some form of staff conference and later are incorporated into the case file of the patient, where they become a permanent part of his record. Although the psychological test report represents the final summary of the psychologist's clinical examination and is viewed as his unique contribution to the total diagnostic study of the patient, comparatively little discussion has appeared in the literature concerning this important activity. Furthermore, while a few accounts concerning the nature and form of psychological reports have appeared [1, 2, 3], critical evaluations of current report writing have been negligible.[2] Because of this lack in a significant area of the psychologist's professional work, it was deemed worth while to report on the attempts of one clinic to evaluate critically the psychological test reports used in the daily activities of the clinic.[3]

The Present Study

The need for the present study was recognized as a result of the regular interdisciplinary staff discussions at the clinic. In the course of these meetings, various dissatisfactions with the psychological test report were voiced. The most important criticisms dealt with the meaningfulness of the psychological report and its general usefulness. As a result of these staff discussions, the psychologists were stimulated to explore the problem at greater length, and a more formal investigation was undertaken in the hope that a better evaluation of the problem and some specific guides toward improving the psychological reports might be obtained.

Procedure

The procedure eventually decided upon was to select a one-month period far enough in the past to avoid ready identification of the author, but recent enough to be representative of current psychological reports. All the reports written during this month, 16 in number, were stripped of all identifying data and constitute the material upon which the present study is based.

The reports were then numbered from 1 to 16. The numbers were placed in a hat and one of them was selected. Copies of the report bearing this number were then typed and distributed to every full-time professional staff person in the clinic. In addition, the remaining 15 reports were retyped and distributed in a random fashion to the clinic staff. In this manner, each staff person was given two psychological reports in a blank envelope and asked to evaluate each of the reports and assign a letter grade to it. The evaluations were to be done anonymously and returned to a secretary. The only identification on the envelope related to the professional group to which the staff member belonged. In this way we were able to compare the evaluations of psychiatrists, psychologists, and social workers as a group. Specific written instructions were included as follows:

---

[1] From the VA Hospital, Downey, Illinois and the VA Regional Office, Chicago, Illinois.

[2] The first book on writing clinical reports [1] was published as this manuscript was being prepared.

[3] The project reported here was a cooperative one in which all staff members of the VA Mental Hygiene Clinic, Chicago, Illinois, participated. The authors wish to acknowledge this participation and to thank especially Drs. D. G. Wright, John G. Watkins, and Miss Clara Weimer for their aid in completing the project.

> Your cooperation and participation is requested in this survey (anonymously). The Psychology Department believes that by securing independent evaluations from the various members of the staff regarding psychological reports, it will be possible to incorporate many valuable suggestions to the end of better service in the diagnostic area.
>
> In order that there may be specific remarks with which we can work, each participant in this survey is being supplied with two sample reports, one of which is being evaluated by everyone. There are wide margins where any notations or remarks may be made regarding the adjacent paragraphs, phrases, etc. The remarks may cover any aspect, from theory to grammar. In fact, feel free to cover any range of comment you wish.
>
> By way of explanation, the reports used have been taken from consecutive referrals over a one-month period and have been written by all members of the department who were examining during that period. The numbers on the reports are merely for purposes of tracing them back to the original testing material.
>
> Please give a letter rating at the top of the first page to indicate the over-all quality of the report, e.g., A, excellent; B, good; C, fair; D, poor; F, very poor.

The reports were then given to 9 psychologists, 13 psychiatrists, and 11 psychiatric social workers. All of the staff with the exception of one psychiatrist returned the reports with their written evaluations. Thus, there were 32 sets of evaluations. No attempt was made to ask specific questions or to have the participants rate the reports on selected variables. Rather, we were interested in the spontaneous and free expression of opinion on any aspect of the psychological test reports that might stimulate response, and wanted to leave the matter as unstructured as possible. Although this procedure makes the analysis of the data more difficult, such an approach was deemed advisable in terms of the preliminary nature of the investigation.

Once the evaluations were collected, the writers were selected as a research team to analyze the data. Each of the three authors studied the common report independently; he recorded the comments and tried to order the material in some meaningful manner. Next, several meetings were held to discuss the preliminary results obtained and to arrive at a uniform method of recording the responses to the test report. A check list consisting of six categories with 24 items was then prepared to use in reordering the data in a uniform and meaningful framework. The common report was again reappraised independently by all three investigators in terms of this guide. It was decided at that time also to record a specific comment or criticism only once for each person. For example, if a participant made the same criticism of one report a number of times, it was recorded only once. This step was judged necessary to compensate for the variability among the written comments of individual staff members and as a check against distorting the significance of specific comments. Thus, if one person made the same criticism 10 times, it would receive the same weight as a general criticism made once by each of 10 different individuals. As a consequence, the unit used was the number of individuals making the comment rather than the actual number of times a criticism was made.

After the second independent appraisal of the responses to the common report, the writers reassembled to discuss their findings and plan for the final recording of the data. The final recording and tabulation of responses to all reports was then made by the writers as a team.

Results

The letter grades given to the common report are summarized in Table 1. In general,

Table 1

Grades Given to Psychological Test Reports

Grade	Common report		
	Psychol.	Psychiat.	Soc. wkrs.
A	1	1	5
B	3	7	3
C	5	1	3
D & below		2	
Total	9	11*	11
Grade	**Other reports**		
	Psychol.	Psychiat.	Soc. wkrs.
A	2	0	0
B	2	1	8
C	4	7	2
D & below	1	4	1
Total	9	12*	11

* One psychiatrist submitted written comment but omitted a grade for the common report.

the report was judged to be satisfactory or better. Social workers were inclined to rate the report much higher than representatives of the other two clinic disciplines. While the median rating of the psychiatrists was higher than that of the psychologists, the only participants who gave the report a rating below C were psychiatrists. This finding may indicate a greater variability in the evaluative attitudes toward psychological test reports held by psychiatrists.

The grades given for the other reports differ in some ways from the distribution obtained for the common report. In general, the social workers and psychiatrists gave lower grades to these reports than to the common report, whereas the psychologists' grades were essentially similar in both instances. Since there were some differences in these reports, no firm conclusions can be reached, but certain trends are discernible. In terms of all the reports evaluated, the psychologists demonstrated the greatest stability of ratings, the social workers tended to give the most positive evaluations, and the psychiatrists the most critical evaluations. This trend was also apparent in the qualitative comments made for each of the test reports.

As indicated previously, some difficulties were encountered in trying to order the spontaneous comments of the staff with regard to the psychological test reports. We purposely wanted a free expression of opinion. However, since the verbatim responses varied greatly, some way to group the data was necessary. The particular classifications finally used are given in Table 2. Attempts were made to use the actual verbalizations and meanings as subheadings as much as possible to avoid misinterpretations of the data. Undoubtedly some of the categories overlap, and some of the different groupings may in part reflect specific orientations of each professional discipline.

As we have organized the data, the critical comments fall into four major categories: (*a*) lack of behavioral referents or supporting data for statements contained in the reports; (*b*)

Table 2

Evaluative Responses to Psychological Test Reports

Categories of evaluative responses	Common test report				Other test reports			
	Psychol. ($N = 9$)	Psychiat. ($N = 12$)	Soc. wkrs. ($N = 11$)	Total ($N = 32$)	Psychol. ($N = 9$)	Psychiat. ($N = 12$)	Soc. wkrs. ($N = 11$)	Total ($N = 32$)
Lack of supporting data or behavioral referents								
No behavioral referents	7	2	1	10	5	1	3	9
Too speculative	0	6	3	9	2	9	3	14
No supporting test data	0	3	0	3	1	8	3	12
Poor expression								
Use of clichés, jargon, etc.	4	1	1	6	5	7	3	15
Loose use of terms	4	2	0	6	5	5	0	10
Vague and unclear	4	0	0	4	4	9	3	16
Poorly expressed	4	0	0	4	2	2	1	5
Poor organization								
No integrated picture of patient	3	0	0	3	2	1	1	4
Poorly organized	0	0	2	2	3	2	2	7
Miscellaneous								
Inconsistent	3	4	1	8	4	8	0	12
Incorrect use of theory	3	3	1	7	0	3	0	3
Test data not differentiated from other data	1	1	2	4	0	2	0	2
Makes unconscious sound conscious	2	2	0	4	1	0	0	1
Doesn't answer referral problem	0	0	2	2	2	1	1	4
Too long and irrelevant	0	0	0	0	4	4	4	12

general lack of clarity of expression; (*c*) poor organization; and (*d*) miscellaneous observations. The first two categories appear to be the most significant in terms of the qualitative comments received. All of the categories, however, will be discussed briefly.

The psychologists reacted most critically to a lack of specific behavioral referents, i.e., individualized examples of very general statements made in the test reports. For example, if a patient was characterized as anxious, depressed, or immature, they wanted to know how this was manifested behaviorally or what the significance of the trend was for the specific patient. Two excerpts from one of the test reports that exemplify this deficiency are the following:

> He has little ability to delay gratification through the medium of mature fantasy activity and thus resorted to immature, wishful inner living.
>
> There is little hazard of the patient acting out since there is a paucity of inner resources.

These statements, as they stand, defy easy translation into terms readily referable to behavior, irrespective of the professional background brought to them by the reader. Thus, it is not surprising that such expressions in a psychological report come under criticism. One of the basic questions — "does this man have strong impulses to commit acts of violence and, if so, how does he cope with them?" — remains unanswered. In the opinion of the writers, the report properly should include a description of the kinds of provocation which evoke hostile aggressive impulses and the ways in which angry feelings are handled by the patient short of directing destructive actions against himself or others. The reader would then be in a position to evaluate along with the diagnostician the "hazard" of this patient's "acting out."

While the psychologists were concerned with the problem of behavioral referents, the psychiatrists reacted to many general statements of personality dynamics with obvious skepticism. Over half of them felt that statements in the test reports were "too speculative" or lacked supporting data. Typical comments in this general category were the following: "I object to this immediate plunge into interpretations, speculations and conclusions without citing test data highlights for each of the tests administered on which to base opinions." "On what is this speculation based? . . . Data, at least some, are essential to evaluate this speculation." "What evidence is there for this interpretation?" "There is a tendency too often to accept certain hypothetical dynamics as fact and base conclusions on such theoretical conclusions." "Again, suggest at least some basic data as obtained and separate from interpretaton and/or speculation." One final illustration is offered as an example of what we considered a conscientious appraisal of the report:

> This portion gives the impression of being based on test results; however, the following fact does not leave such an impression but suggests instead that test material was used as a take-off point for a flight into the type of speculation that suits the examiner's system of psychological thought. Perhaps the means to avoid such impression might be the use of examples from the test material wherever possible, e.g., "as evidenced by," "as reflected in," etc., to distinguish the demonstrable from the supposed.

There was little question that the psychiatrists judged these kinds of criticisms to be among the most important. Such criticisms were also offered by approximately one fourth of the social workers. The reasons for their skepticism are not readily discernible, but it is worth "speculating" about them. It may be that the reports appeared to be too inclusive and that they doubted that such complete portrayals of personality could be obtained solely from test data. Another consideration may be the tone of finality evident in many reports which also tends to stir negative reactions. Undoubtedly, other more obvious factors are also involved. Very evident inconsistencies, statements which contradict other data, and similar factors may have tended to undermine confidence in the reports and to cause one to view them with suspicion. Whatever the factors may have been, the net result was to limit the acceptance of the test reports.

The criticism of lack of supporting data was also accepted by many psychologists with some reservations. It is often difficult, if not impossible, to present evidence succinctly since it comes from a variety of test data. Moreover, psychologists may have some legitimate doubts that members of allied professions are equipped to interpret the evidence which is presented. The solution of this problem may depend on a general improvement in psychological reporting with perhaps some supporting evidence for certain kinds of inferences.

The second major category of critical responses included comments pertaining to poor and careless expression. The following are representative comments: "vague, unclear, loose use of terms, pure jargon, don't understand, poorly expressed, etc." While this category conceivably overlaps with statements in the first major subdivision, in general the emphasis in the second grouping was on lack of clarity in presentation, poor communication, and the use of technical jargon.

There was no lack of examples of vagueness, obscurity, and overdependence upon clichés in the reports examined. The following are quotations in which the "sins" of the psychologists are self-evident:

1. Hostile trends are strong. He employs reaction formation as a means of control but there are strong residual tensions which strive for access to the motor sphere.

2. He has little internal resources remaining, along with heightened dysphoria and anxiety.

3. There is evidence of some dissociative disturbance of genital sexuality.

4. He attempts to be orderly, to draw closure on his experiences which tends to limit spontaneous activity. Hysterical qualities characterize his behavior in that he has regressed psychosexually, is emotionally over-active on a superficial level and has repressed most unacceptable impulses.

In the third category, "poor organization," the majority of critics referred to the following: (*a*) a failure on the part of the diagnostician to discuss related findings in the same paragraph or section of the report; and (*b*) failure to construct the report in such a way that various aspects of the personality discussed separately were finally integrated into a unified picture of the functioning of the patient's personality. Psychologists, primarily, were critical of this latter aspect of the reports.

In the final category of miscellaneous items we included: (*a*) observations of contradictory statements which occurred not only within the same report but occasionally within a single sentence; (*b*) instances where the diagnostician clearly did not understand the personality theory that he was employing as a guide for his interpretations; (*c*) comments noting the inclusion of what was obviously social history data as a test finding; (*d*) statements concerning the length of reports and the failure to answer the referral problem; and (*e*) criticisms about "making the unconscious sound conscious." The latter referred to personality descriptions that sounded as if the individual patient consciously planned his behavior in the manner described, e.g., "since he could find no mature means of gratifying his needs, he resorted to somatization as a means of obtaining gratification."

The matter of inconsistency was not limited to any specific area, but in some cases was a reflection of an uncritical use of test terminology and clinical concepts. An example of the latter is evident in the following comment: "How can he [the patient] be 'affectively impoverished' and at the same time show 'heightened dysphoria' and anxiety — do not the latter indicate capacity for affect?"

The critical comments on the length of reports were made by approximately one third of the staff in reference to reports other than the common one. Whereas the common report was two pages in length, the other reports varied from three to five pages, and this fact may account for the relatively lower ratings given to these reports. These findings would imply that a concise report more specifically directed to the referral problem is much better received than a long rambling one that attempts to present an all-inclusive personality description.

Finally, it should be mentioned that there were, as might be expected, scattered criticisms of grammar, "English," and sentence structure. These, however, were sufficiently limited in number as to warrant bare mention here. Also, a number of favorable comments were made, largely by the social workers. This finding seems consistent with their tendency to be less critical and in general to give higher evaluations to the reports. The positive comments were such as to add weight to the categories of criticism shown in Table 2; that is, reports were viewed favorably for being "concise," "clear," "well written," and directed to answering the referral question. The term "good" was used relatively often with no specific reason given for its use. Perhaps one quotation might help indicate some of the bases for favorable reaction.

> I feel that this is an excellent type of report principally because the language is *simple*, i.e., nontechnical, and the construction is direct and to the point. The "lay-out" too is good.

Summary and Conclusions

A one month's sample of psychological test reports was distributed to all professional staff members in the mental hygiene clinic for their critical evaluation. All identifying data on the reports were removed, and the evaluations were done anonymously. The authors of the study first analyzed the evaluative comments separately and worked out a list of categories which would facilitate the ordering of the data and yet retain the emphases given by the individual staff members. Later the authors worked together on the final tabulation and appraisal of the data.

Some differences between the three professional disciplines were noted. In general, the psychiatrists tended to be most critical of the reports, while the social workers were least critical. The psychiatrists most frequently criticized the reports for being too speculative, omitting supporting data, being vague, unclear, and inconsistent. The psychologists were primarily critical of the lack of behavioral referents for generalized statements made in the reports. They were also sensitive to the use of clichés, technical jargon, and the lack of clarity in several of the reports. In general, the main criticisms of the reports were related to the lack of supporting data or behavioral referents and to a lack of clarity and concise communication.

On the basis of this pilot study, it appears that if psychological test reports are to be more useful they should be as concise as possible, avoid generalities and the use of technical jargon, give supporting data for certain kinds of inferences, avoid "speculative dynamics," and try to give a portrayal of a specific individual using behavioral terms wherever possible. Partly as a result of this study, the staff psychologists at the Chicago VA Mental Hygiene Clinic are revising their style of reporting psychological test data.

Received November 5, 1953.

References

1. Hammond, K. R., & Allen, J. M., Jr. *Writing clinical reports.* New York: Prentice-Hall, 1953.
2. Holzberg, J. D., Alessi, S. L., & Wexler, M. Psychological case reporting at psychiatric staff conferences. *J. consult. Psychol.*, 1951, **15**, 425-429.
3. Sargent, H. D. Psychological test reporting: an experiment in communication. *Bull. Menninger Clin.*, 1951, **15**, 175-186.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

Some Common Misconceptions About Neuroticism[1]

Harrison G. Gough
University of California, Berkeley

A respectable science of psychopathology needs to be based upon, among other things, a thorough and accurate knowledge of what particular beliefs, values, self definitions, attitudes, and self conceptions are characteristic of persons diagnosable as maladjusted. Theories about process, about the "self," about motivations, and about therapeutical techniques can hardly be evolved without some minimum pre-establishment of "facts" in this realm. That is, for example, if one is to develop a theory to account for the contention that most neurotics say that they get a "raw deal" from life, then we must first make sure that most neurotics *do,* in fact, express this belief.

It is rather surprising, if one reviews the psychological literature of the past 20 years, to discover how little has been done to determine the characteristic and differentiating beliefs and attitudes of persons classified with respect to their psychological adjustment. We have some general notions, for example, that neurotics feel misunderstood, that they often have bad dreams, that they often expect to fail in what they do, that they desire more sympathy than they get, that they find it hard to concentrate, and so forth. Yet, in spite of the diffusion of ideas such as these, there are only a few studies which have actually asked patients for responses on these issues and then compared the frequency of their agreements with those of control groups.

The few studies done, such as those of Page, Landis, Katz, and Zubin [6, 7, 10], and Hathaway and McKinley [4, 5, 8], have revealed findings strikingly discrepant from the pervasive stereotypes of both professional and lay persons. The clinical stereotype about what patients would say and think appeared, in both series of studies, to be so wide of the mark as to lead not only to irrelevant designations, but to erroneous ones. For example, in one paper of Page, Landis, and Katz [10], a 50-item scale for schizophrenia based on expert psychiatric judgment actually yielded *higher* scores for a control sample of nonhospitalized adults than it did for a group of diagnosed schizophrenics. Clearly something is amiss in the clinical conceptions of verbal behavior in psychopathology if errors of this magnitude can appear.

None of the studies mentioned, however, was primarily devoted to specifying and outlining the details of the conventional stereotypes about neuroticism, although all of them revealed something of the grossness of the errors of prevailing opinion. The decision was made, accordingly, to attempt an extensive survey of specific stereotypes about neuroticism and then, by comparing the stereotype with the observed reactions of diagnosed patients, to delineate some of the discrepancies between opinion and fact. In making such a search for discrepancies, the broader the sampling of beliefs and attitudes the better. The Minnesota Multiple Personality Inventory [5], by reason of its systematic coverage of many areas of self definition and personal values, is exceptionally well suited to the task at hand, and was therefore selected as the research instrument.

Method of Analysis

Seven samples of subjects were obtained in order to conduct the analysis. Four of these

---

[1] This project was conducted under a research grant from the National Institute of Mental Health, National Institutes of Health, U. S. Public Health Service.

Table 1
Sample Items Revealing Erroneous Conceptions about Neuroticism

MMPI item number	Percentage saying "True"						
	Dissemblers			Patients			
	Professional workers ($N = 11$)	Male students ($N = 50$)	Female students ($N = 50$)	Langley Porter males ($N = 39$)	Oakland VA Clinic males ($N = 46$)	Army males ($N = 40$)	Langley Porter females ($N = 51$)
16. I am sure I get a raw deal from life.	64	82	90	8	7	28	12
35. If people had not had it in for me I would have been much more successful.	36	68	86	0	4	2	0
88. I usually feel that life is worth while.	36	28	10	79	74	60	59
212. My people treat me more like a child than a grown-up.	45	76	76	15	20	2	22
257. I usually expect to succeed in things I do.	27	36	10	72	76	88	67
459. I have one or more bad habits which are so strong that it is no use in fighting against them.	73	80	88	8	13	15	12

samples consisted of patients psychiatrically diagnosed as "psychoneurotic," and the remaining three were experimental samples requested to take the role of a patient experiencing a psychoneurotic reaction and to answer the test items accordingly. The samples were constituted as follows:

- I. Males
  - A. Patients
    - 1. Thirty-nine patients from the University of California Langley Porter Psychiatric Clinic.[2]
    - 2. Forty-six patients from the Oakland VA Mental Hygiene Clinic.[3]
    - 3. Forty patients from a U.S. Army station hospital.[4]
  - B. Dissimulators
    - 1. Fifty students from a University of California, Berkeley, class in abnormal psychology.[5]
    - 2. Eleven professional workers from the Army hospital.
- II. Females
  - A. Patients.
    - 1. Fifty-one patients from the Langley Porter Psychiatric Clinic.
  - B. Dissimulators
    - 1. Fifty students from the class in abnormal psychology.

For each of the items in the MMPI, the proportion in each of the above samples replying "True" was calculated, and then three internal comparisons (I. A. 1 and I. A. 2 vs. I.-B.1; I.A. 3 vs. I.B.2; and II.A. 1 vs. II. B.1) were made. All items yielding similar and statistically significant differences in each comparison were selected. These items, then, could be said to embody the fallacious stereotypes of the experimental samples. The entire statistical table is too long to be given here,[6] but a few selected items are listed in Table 1 to indicate the impressive magnitude of error observed. Note, for example, item number 35: "If people had not had it in for me I would have been much more successful." In the stereotype about what patients will say, this sentiment features strongly, and yet, in the actual responses of patients, it is almost nonexistent. In the light of this particular finding what is one to make of a theory which attempts to account for the "fact" that psychoneurotic patients tend to harbor such feelings, and to express them on inquiry?

---

[2] These records were very kindly furnished by Dr. R. E. Harris.

[3] These records were very kindly furnished by Dr. G. S. Welsh.

[4] These records were drawn from an earlier study by the writer [see 1, 2].

[5] These records were very kindly furnished by Dr. T. R. Sarbin.

[6] A complete version of this table, giving all discriminating items and the proportion of each sample answering "True" has been filed with the American Documentation Institute. Order Document No. 3631 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.00 for microfilm or $1.00 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Altogether, 74 items were identified which met the criterion of significance.[7] These items, and the direction of response given more frequently by the simulators, are listed below:

10. There seems to be a lump in my throat much of the time. (True) 14. I have diarrhea once a month or more. (True) 16. I am sure I get a raw deal from life. (True) 19. When I take a new job, I like to be tipped off on who should be gotten next to. (True) 23. I am troubled by attacks of nausea and vomiting. (True) 24. No one seems to understand me. (True) 29. I am bothered by acid stomach several times a week. (True) 31. I have nightmares every few nights. (True) 35. If people had not had it in for me I would have been much more successful. (True) 42. My family does not like the work I have chosen (or the work I intend to choose for my life work). (True)

44. Much of the time my head seems to hurt all over. (True) 47. Once a week or oftener I feel suddenly hot all over, without apparent cause. (True) 50. My soul sometimes leaves my body. (True) 53. A minister can cure disease by praying and putting his hand on your head. (True) 68. I hardly ever feel pain in the back of the neck. (False) 73. I am an important person. (True) 83. Any man who is able and willing to work hard has a good chance of succeeding. (False) 88. I usually feel that life is worth while. (False) 93. I think most people would lie to get ahead. (True) 96. I have very few quarrels with members of my family. (False)

97. At times I have a strong urge to do something harmful or shocking. (True) 104. I don't seem to care what happens to me. (True) 125. I have a great deal of stomach trouble. (True) 137. I believe that my home life is as pleasant as that of most people I know. (False) 179. I am worried about sex matters. (True) 206. I am very religious (more than most people). (True) 207. I enjoy many different kinds of play and recreation. (False) 210. Everything tastes the same. (True) 211. I can sleep during the day but not at night. (True) 212. My people treat me more like a child than a grown-up. (True)

216. There is very little love and companionship in my family as compared to other homes. (True) 226. Some of my family have habits that bother and annoy me very much. (True) 241. I dream frequently about things that are best kept to myself. (True) 246. My neck spots with red often. (True) 247. I have reason for feeling jealous of one or more members of my family. (True) 257. I usually expect to succeed in things I do. (False) 297. I wish I were not bothered by thoughts about sex. (True) 303. I am so touchy on some subjects that I can't talk about them. (True) 306. I get all the sympathy I should. (False) 320. Many of my dreams are about sex matters. (True)

325. The things that some of my family have done have frightened me. (True) 328. I find it hard to keep my mind on a task or job. (True) 341. At times I hear so well it bothers me. (True) 344. Often I cross the street in order not to meet someone I see. (True) 352. I have been afraid of things or people that I knew could not hurt me. (True) 360. Almost every day something happens to frighten me. (True) 375. When I am feeling very happy and active, someone who is blue or low will spoil it all. (True) 388. I am afraid to be alone in the dark. (True) 405. I have no trouble swallowing. (False) 419. I played hooky from school quite often as [illegible] youngster. (True)

422. I have felt embarrassed over the type of work that one or more members of my family have done. (True) 433. I used to have imaginary companions. (True) 438. There are certain people whom I dislike so much that I am inwardly pleased when they are catching it for something they have done. (True) 443. I am apt to pass up something I want to do because others feel that I am not going about it in the right way. (True) 453. When I was a child I didn't care to be a member of a crowd or gang. (True) 458. The man who had most to do with me when I was a child (such as my father, stepfather, etc.) was very strict with me. (True) 459. I have one or more bad habits which are so strong that it is no use in fighting against them. (True) 466. Except by a doctor's orders I never take drugs or sleeping powders. (False) 471. In school my marks in deportment were quite regularly bad. (True) 475. When I am cornered I tell that portion of the truth which is not likely to hurt me. (True)

476. I am a special agent of God. (True) 480. I am often afraid of the dark. (True) 481. I can remember "playing sick" to get out of something. (True) 485. When a man is with a woman he is usually thinking about things related to her sex. (True) 518. I have often felt guilty because I have pretended to feel more sorry about something than I really was. (True) 519. There is something wrong with my sex organs. (True) 524. I am not

---

[7] The 44 MMPI items revealing the greatest differentiations are included (by permission of the copyright owners) as a "dissimulation scale" in the writer's California Psychological Inventory, copyright 1951. This inventory is one designed to assess variables having broad social and interactional relevance, rather than those pertaining to special medical, psychiatric, or vocational categories. Scales for variables such as social responsibility, dominance and leadership, intellectual efficiency, and academic motivation are included. [See, for example, 3.]

afraid of picking up a disease or germs from door knobs. (False) 525. I am made nervous by certain animals. (True) 528. I blush no more often than others. (False) 535. My mouth feels dry almost all the time. (True)

541. My skin seems to be unusually sensitive to touch. (True) 543. Several times a week I feel as if something dreadful is about to happen. (True) 545. Sometimes I have the same dream over and over. (True) 565. I feel like jumping off when I am on a high place. (True)

The similarity of these items to a conventional neuroticism questionnaire is obvious. Most observers, professionally trained or not, would probably identify the items at once as pertaining to maladjustment. Nevertheless, the inference seems justified that they do *not* pertain to neuroticism and maladjustment, but rather to the prevailing *stereotypes* about neuroticism.[8]

For each of the specific instances listed, neurotic patients answer in precisely the opposite direction from what is predicted for them. They do not, in other words, respond "True" to questionnaire items like "No one seems to understand me," "I don't seem to care what happens to me," and "I have one or more bad habits which are so strong that it is no use in fighting against them," and they do respond "True" to items like "I get all the sympathy I should, " and "I usually expect to succeed in things that I do." Even such a short listing as this should serve to highlight the dangers that are involved in any a priori decisions about how patients will answer questions such as these. Indeed, as will be shown below, the way to identify psychoneurotic predispositions as they are manifested on the above list would not be to accept current theory in scoring each item, but rather to reverse it in every instance and score each item backwards.

In order to simplify and condense the list of 74 items, the writer has attempted a subjective clustering. The groupings which were derived are as follows:

1. Numerous physical complaints and dysfunctions.
2. Feelings of victimization, injustice, and misunderstanding; lack of hope for and faith in the future.
3. Exaggerated irritability, petulance, tenseness, fear, and anxiety.
4. Lack of independence and self-sufficiency; inability to face issues squarely.
5. Dissatisfaction with family background and childhood.
6. Sexual conflicts and preoccupations.
7. Bizarre, eccentric ideas; excessive religiosity.

These clusters, then, summarize to some extent the kinds of emphases which are given too much weight in the typical stereotypes about the expressed beliefs and attitudes of maladjusted persons.

Cross-Validating Samples

The separate item discriminations achieved in the analyses were sufficiently striking to suggest trying them as a scale for dissimulation. For this purpose a number of new samples were obtained. The results are indicated in Table 2. Of the 13 samples in Table 2, only one, the 11 Army dissemblers, was used in the item analyses just reported. All of the other samples include cross-validating cases only, so that the results are probably similar to what would be anticipated in most subsequent applications of these items as a scale.

The analysis of variance over 13 separate samples yielded an F ratio of 374.17, significant well beyond the .01 probability point. The clinical samples and the student samples both average near the lower end of the scale, with the dissimulators all scoring much higher.

If the eight clinical samples are pooled in a composite ($N = 915$), a mean of 15.94 and a standard deviation of 9.99 are obtained. For the student samples the combined figures ($N = 507$) are 15.88 and 7.90. The difference between the means is .06, and the CR, 0.125. This is an important finding, for it indicates that patients diagnosed as psychoneurotic actually achieve mean scores on this scale nearly identical with those attained by a large nonhospitalized sample. Patients not only do not respond to the questions in the manner indicated by usual conceptions, but are actually indistinguishable from normals in reacting to the set of 74 statements. Furthermore, both patients and control samples score considerably below the mean achieved by subjects asked to simulate a psychoneurotic reaction.

The distributions of scores for the three con-

---

[8] The findings also tend to substantiate the viewpoint that structured personality tests may contain subtle and dynamic features over and above what the clinician may recognize in a casual inspection [see 9].

Table 2
Means and Standard Deviations on the 74-Item Dissimulation Scale for the Samples Indicated

Sample	*N*	*M*	*SD*
A. Dissemblers			
1. Professional workers	11	46.82	11.21
2. Students in abnormal psychology	302	53.96	11.52
3. Students in personality in society and culture	41	57.34	11.90
B. Clinical cases			
1. Army psychiatric patients	250	18.87	10.94
2. Ft. Snelling VA Hospital males*	99	16.91	11.45
3. Oakland VA Hospital males**	103	13.40	10.20
4. San Francisco VA Mental Hygiene Clinic males**	62	17.45	9.33
5. Langley Porter Psychiatric Clinic males†	100	15.05	9.21
6. Langley Porter Psychiatric Clinic females†	100	16.25	9.58
7. University of Minnesota Psychiatric Division males‡	101	12.31	7.02
8. University of Minnesota Psychiatric Division females‡	100	13.59	6.86
C. Nonclinical samples			
1. High school freshmen	223	16.89	8.32
2. High school seniors	284	15.08	7.48

* These records were very kindly furnished by Dr. William Hales.
** These records were very kindly furnished by Dr. G. S. Welsh.
† These records were very kindly furnished by Dr. R. E. Harris.
‡ These records were very kindly furnished by Dr. S. R. Hathaway.

solidated samples—dissemblers, clinical cases, and students—are presented in Table 3.

The *F* ratio for the threefold breakdown represented in Table 3 was 2,141.04, a clearly significant value. The screening efficiency of the *Ds* scale for these cross-validation samples (all were cross-validating cases except for the 11 Army dissimulators) can be exampled by a cutting score of 35. Ninety-three per cent of the dissemblers scored at or above this point, whereas only 6 per cent of the clinical cases and 2 per cent of the students did so.

Table 3
Proportion of Cases Falling in the Scoring Categories Indicated

Scores	Dissemblers (*N*=354)	Clinical cases (*N*=915)	Students (*N*=507)
70-74	4.2		
65-69	15.2		
60-64	19.4		
55-59	16.9		
50-54	14.7	0.4	
45-49	11.0	0.8	0.2
40-44	7.3	1.8	0.6
35-39	4.2	2.6	1.2
30-34	3.7	5.3	4.5
25-29	2.0	7.2	7.5
20-24	0.6	12.5	14.8
15-19	0.3	17.5	23.2
10-14	0.3	21.5	24.6
5- 9		19.8	19.3
0- 4		10.1	3.9
M	54.13	15.94	15.88
SD	11.69	9.99	7.90

The results obtained in treating the differentiating items as a *scale* further underscore the erroneousness of the particular aspects of the clinical stereotypes of neuroticism they embody. The opinions and sentiments contained in the items are not expressed, or more accurately, assented to, with any *more* frequency by diagnosed psychoneurotic patients than by control samples, and, furthermore, are agreed with much *less* frequently than the conventional conceptions of neuroticism would presage.

Summary

An investigation of typical conceptions of neuroticism was conducted by having subjects take the role of a psychoneurotic patient in responding to a personality inventory. Significant discrepancies between what diagnosed patients did, in fact, report on this inventory and the stereotypes given by the simulators were discovered.

The argument was advanced that errors of the magnitude observed suggested a considerable degree of misinformation in the prevailing conceptions about neuroticism.

Finally, it was possible, by consolidating the items individually tapping these misconceptions, to develop a measuring instrument yielding similar scores for both patient and control samples, and differentiating each from samples based on dissembled records.

Received November 16, 1953.

References

1. Gough, H. G. Simulated patterns on the MMPI. *J. abnorm. soc. Psychol.*, 1947, **42**, 215-225.
2. Gough, H. G. The *F* minus *K* dissimulation index on the MMPI. *J. consult. Psychol.*, 1950, **14**, 408-413.
3. Gough, H. G. The construction of a personality scale to predict academic achievement. *J. appl. Psychol.*, 1953, **37**, 361-366.
4. Hathaway, S. R., & McKinley, J. C. A Multiphasic personality schedule: I. Construction of the schedule. *J. Psychol.*, 1940, **10**, 249-254.
5. Hathaway, S. R., & McKinley, J. C. *The Minnesota Multiphasic Personality Inventory.* Minneapolis: Univer. of Minnesota Press, 1943.
6. Landis, C., & Katz, S. E. The validity of certain questions which purport to measure neurotic tendencies. *J. appl. Psychol.*, 1934, **18**, 343-356.
7. Landis, C., Zubin, J., & Katz, S. E. Empirical evaluation of three personality adjustment inventories. *J. educ. Psychol.*, 1935, **26**, 321-330.
8. McKinley, J. C., & Hathaway, S. R. A multiphasic personality inventory: V. Hysteria, hypomania, and psychopathic deviate. *J. appl. Psychol.*, 1944, **28**, 153-174.
9. Meehl, P. E. The dynamics of structured personality tests. *J. clin. Psychol.*, 1945, **1**, 296-303.
10. Page, J., Landis, C., & Katz, S. E. Schizophrenic traits in the functional psychoses and in normal individuals. *Amer. J. Psychiat.*, 1934, **90**, 1213-1225.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

Factors Descriptive of Chronic Schizophrenics Selected for the Operation of Prefrontal Lobotomy

Maurice Lorr, Richard L. Jenkins, and James Q. Holsopple
Veterans Administration, Washington, D. C.

This study was undertaken in the hope of contributing to a simpler and conceptually more satisfactory differentiation of the major schizophrenic processes. To this end a set of scores descriptive of a group of chronic schizophrenics, chosen for the operation of prefrontal lobotomy, and their matched controls were subjected to a multiple-factor analysis.

Procedure

The data consisted of clinical judgments recorded on the Northport Record [4, 7], a schedule of 81 brief rating scales for describing the characteristic behavior and symptoms of psychotic patients. The ratings were made by psychiatrists and psychologists on 153 schizophrenic patients at six Veterans Administration hospitals which are participating in a study to define the nature and extent of changes that occur as a result of the operation of prefrontal lobotomy [3]. All ratings were obtained prior to operation on patients later lobotomized or on patients selected to serve as controls. Since no patient had been mentally ill for less than 4 years, the data represent a group of patients whose psychoses had become chronic.

The 11 scores included in the study were based on a factor analysis of ratings of 423 hospitalized patients on 55 of the scale variables included in the Northport Record [5]. Each score consisted of the sum of the ratings received by the patient on a group of scales best defining a factor. For example, the paranoid projection factor *C* consisted of the sum of the ratings on scales descriptive of ideas of reference, delusions of persecution, delusions of influence, and frequency of assaults. There were relatively few scales common to two or more factors. Factor *A*, depressed retarded versus manic excited, and factor *E*, melancholy agitation, had 4 scales in common. A few of the factors had two scales in common. Of the 55 possible relations between factors there is no overlap in 34.

The 11 factors scored have been named as follows:

A. Retarded depression versus manic excitement

B. Compliance versus resistiveness

C. Paranoid projection

D. Activity level

E. Melancholy agitation

F. Perceptual distortion

G. Motor disturbances

H. Submissiveness versus belligerence

I. Aggressive social interest versus social withdrawal

J. Self-depreciation versus grandiose expansiveness

K. Conceptual disorganization

The Pearson correlations computed between the 11 factor scores are presented in Table 1. The correlational matrix was subjected to a centroid analysis with centroid estimates of the communalities in the diagonals, and 3 factors were extracted. The median absolute third factor residual was .06 as compared to the standard error of .08 of a zero correlation on a population of this size. The orthogonal factors and the rotated oblique factors are presented in Table 2. The cosines between the reference vectors are as follows: Alpha versus Beta, .07; Alpha versus Gamma, .51; Beta versus Gamma, .05. The transformation matrix is shown in Table 3.

Table 1
Intercorrelations of Raw Scores on 11 Factors of 153 Schizophrenic Patients

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
A	−.13									
B	.01	.03								
C	.16	.02	.02							
D	.53	−.28	.06	.04						
E	−.04	.39	.46	.10	−.22					
F	−.11	.42	.17	.06	−.27	.02				
G	.32	.50	.33	.23	.25	.16	−.24			
H	−.54	.39	−.06	−.23	−.36	.01	.35	.71		
I	.29	.40	.14	.22	−.05	.27	.12	.57	−.01	
J	−.28	.53	.28	.00	−.47	.71	.63	−.08	.40	.11

The Findings

The first factor, Alpha, is defined by withdrawal (*I*), conceptual disorganization (*K*), motor disturbances (*G*), absence of melancholy agitation (*E*), the schizoid elements of depressive retardation (*A*), and resistiveness (*B*). A patient with high scores on Alpha exhibits manneristic movements and postures, lacks tension or anxiety, appears apathetic or lacking in feeling. Typically he is mute or taciturn, and when he responds his speech tends to be irrelevant, incoherent, and loaded with stereotyped language. He is also likely to be disoriented for self and others. On the ward he will sit isolated, mute, and indifferent to the activities around him. The parameter underlying Alpha appears to be one of apathetic withdrawal with motor disturbances and conceptual disorganization.

Table 2
Second-Order Centroid Matrix F and Oblique Matrix V

Factors		F				V		
		I	II	III	h^2_j	α	β	γ
Depression*	*A*	—27	61	18	48	55	11	—07
Resistance	*B*	70	07	27	57	49	36	66
Paranoid distortion	*C*	34	25	—24	23	—02	47	08
Overactivity	*D*	09	30	—07	10	—17	20	06
Apathy*	*E*	—45	50	16	48	60	27	08
Perceptual distortion	*F*	60	26	—55	73	05	85	01
Motor disturbances	*G*	49	—36	09	37	60	18	31
Belligerence	*H*	44	63	62	98	04	14	86
Withdrawal	*I*	61	—39	60	89	85	—06	77
Grandiosity	*J*	42	53	16	48	—05	37	48
Conceptual disorganization	*K*	78	—30	—29	79	64	65	22

* A and E and their descriptions have been reversed in matrix V.

The second-order factor, Gamma, which is negatively related to Alpha, is most sharply defined by factor *H* (submissiveness versus belligerence), factor *I* (withdrawal), factor *B* (compliance versus resistiveness), and factor *J* (self-depreciatory versus grandiosity). At one end Gamma is characterized by overt hostility, irritability, assaultiveness, obscenity and profanity, neglect of personal appearance, isolation from others, resistiveness, and destructiveness. Ideation is marked by ideas of grandeur and boastful conceit. The opposite end of Gamma is characterized by a meek submissiveness, compliance to routine, and self-depreciatory tendencies.

The belligerent, resistive, assaultive behavior of Gamma is often associated with an excited agitation. This would account for the substantial negative relation between Alpha and Gamma although the two also have in common withdrawal and resistive tendencies. It is the type of patient high on Gamma who is frequently a candidate for lobotomy.

Table 3
Transformation Matrix

	A	*B*	*C*
I	.66	.72	.63
II	—.69	.32	.18
III	.29	—.62	.75

The second-order factor, Beta, has its highest correlations from perceptual distortion (*F*), conceptual disorganization (*K*), and paranoid projection (*C*). A patient scoring high on this factor is characterized by auditory and visual hallucinations, ideas of reference, and delusions of persecution and influence. He is likely to be irrelevant or incoherent in his speech, and disoriented as to who he is or who others are around him.

Schizophrenic patients considered for lobotomy may, in the foregoing terms, be said to show three distinguishable modes of reaction. They withdraw (Alpha), they distort in perception and thought the world they encounter around them (Beta), and /or they fight (Gamma). If the withdrawal factor Alpha is pronounced and the disorganization of thinking factor Beta is not obvious, the patient is likely to be diagnosed as showing catatonic withdrawal. If disorganization (Beta) is pronounced, the patient will tend to be diagnosed as hebephrenic. If the progress of the disorganization is slowed by the false solution of paranoid projection, the patient will be categorized as a paranoid schizophrenic. If an assaultive belligerence (Gamma) is marked and associated with an agitation (negative Alpha) the patient will be regarded as a good candidate for lobotomy.

Discussion

The three second-order factors identified here will now be compared with those identified by Degan [1], and by Lorr, Jenkins, and O'Connor [5]. Since the patient groups sampled differed considerably in these studies and the scales used were different, we can expect only a general correspondence in the factors identified. Nevertheless, it is of interest to see how much these second-order factors have in common.

Degan utilized data collected by Father T. V. Moore [6] on 367 hospitalized patients diagnosed as functional psychotics. He identified nine first-order factors and four in the second-order domain. Table 4 presents the projections of his first-order factors on the second order. In the study by Lorr *et al.* [5], 55 ratings descriptive of 423 hospitalized psychotic patients were factored. The 11 first-order factors were analyzed and three second-order factors were identified (Table 5).

Table 4

Degan's Second-Order Rotated Oblique Matrix

First-order factors		Second-order factors			
		W	*X*	*Y*	*Z*
Hyperprojection	*A*	−09	40	−52	−01
Depression	*B*	−06	−13	51	−42
Manic hyperexcitability	*C*	56	13	−03	00
Catatonia	*D*	−08	−03	−01	71
Schizophrenic dissociation	*E*	−01	68	01	03
Traumatic hysteria	*F*	01	45	−20	35
Hyperirritability	*G*	61	−01	−37	33
Deterioration	*H*	22	19	17	32
Neurasthenia	*I*	−26	10	63	−03

Degan's second-order factor *Z*, interpreted as catatonic schizophrenia, is most similar to factor Alpha. Factor *Z*, in terms of the original variables, is defined by mutism, refusal of food, stereotypism of attitude, negativism, giggling, and absence of an agitated depression. To a lesser degree, it is characterized by disorientation for time and space, irritability, tantrums, and loss of finer sensibilities. Only the irritability elements of factor *Z* are lacking in our Alpha. It should be noted that Alpha was identified in a group of patients diagnosed as schizophrenic and sufficiently chronic to be considered for lobotomy, while factor *Z* was derived from a mixed psychotic population. Furthermore, in the population from which Alpha was drawn, assaultiveness, hostility, and irritability were sufficiently common as to define factor Gamma (belligerence). It is for these reasons that irritability has its projections on Gamma rather than Alpha.

Table 5

Second-Order Oblique Factor Matrix Based on Intercorrelations between Primaries (Lorr, Jenkins, O'Connor [5])

Factors		*AA*	*BB*	*CC*
Depression	*A*	47	25	34
Resistiveness	*B*	54	06	22
Paranoid distortion	*C*	−08	39	−03
Overactivity	*D*	−21	−42	01
Apathy	*E*	−03	00	54
Perceptual distortion	*F*	09	34	02
Motor disturbances	*G*	56	01	00
Belligerence	*H*	58	21	03
Withdrawal	*I*	45	04	68
Grandiose expansiveness	*J*	−20	−10	26
Conceptual disorganization	*K*	−29	03	13

Factor *AA* (Table 5) is characterized by resistiveness, frequent assaults, manneristic movements and postures, disorientation for time, place, self, and people, a complete lack of interest, and muteness. Alpha has in common with *AA* manneristic movements and postures, disorientation, and withdrawal of interest, but it lacks the assaultiveness of *AA*. In addition, Alpha is characterized by a conceptual disorganization missing in *AA*.

Beta has considerable resemblance to Degan's factor X, which is defined principally by stereotypism of words and of actions, talking to voices, appearing "shut-in," disorientation for space and time, hallucination, and bizarre delusions. Only the resistiveness and grandiosity of Beta are absent.

Beta is probably the same as factor BB in Table 5. Factor BB is defined by delusions of influence, ideas of reference, ideas of persecution, auditory and visual hallucinations, and slowed activity. It lacks only the thinking disorganization of Beta.

Since the present study was confined to severe chronic schizophrenics, it is not surprising that primarily affective factors present in the other two studies are missing. In Degan's study, we have a factor of manic excitement (W) and a bipolar factor of paranoid versus depressive psychosis (Y), i.e., of projections versus introjection. Factor CC, on the other hand, is based on an analysis of a population probably having fewer patients with manic-depressive psychosis and more with schizophrenia. As a consequence, factor CC is more schizophrenic than affective. It ranges from an apathetic withdrawal at one end to excitement with grandiose expansiveness at the other. Factors W, Y, and CC do not appear in the present study of severe chronic schizophrenics.

The belligerence factor Gamma is unique to the present study. No doubt it is due to the selection of a sample of patients considered for lobotomy, many of whom were explosively tense and assaultive. Resemblances between the factors are summarized in Table 6.

Table 6
Similarities between Second-Order Factors Identified in Three Factor Analyses

Hypothesized process	Degan's analysis	Study Lorr, Jenkins, O'Connor analysis	Present analysis
Resistive withdrawal	Z	AA	Alpha
Disorganization and distortion of thinking	X	BB	Beta
Manic excitement	W	CC(?)	Not found (schizophrenic patients)
Projection versus introjection	Y		
Belligerence	Not found	Not found	Gamma

Summary

A factorial analysis was made of the intercorrelations between 11 relatively independent scores descriptive of the behavior and psychopathology of 153 chronically ill schizophrenic patients selected for prefrontal lobotomy and their matched controls. Three common factors were extracted and rotated to simple structure. Alpha is interpreted as a factor of apathetic withdrawal with motor disturbances. Beta is a factor of perceptual and thinking distortion which may represent a phase of schizophrenic disorganization. Gamma describes a fighting reaction. The negative relation between Alpha and Gamma suggests a third-order bipolar parameter of apathetic withdrawal versus agitated belligerence.

The first two factors resemble those found in Degan's analysis of Father Moore's data and those isolated in the Lorr, Jenkins, and O'Connor study.

Received November 18, 1953.

References

1. Degan, J. W. Dimensions of functional psychosis. *Psychometr. Monogr.*, 1952, No. 6.
2. Jenkins, R. L. The schizophrenic sequence: withdrawal, disorganization, psychotic reorganization. *Amer. J. Orthopsychiat.*, 1952, **22**, 738-748.
3. Jenkins, R. L., & Holsopple, J. Q. Criteria and experimental design for evaluating results of lobotomy. *Res. Publ. Ass. nerv. ment. Dis.*, 1951, **31**, 319-327.
4. Lorr, M. Multidimensional scales for rating psychiatric patients. Hospital form. *Vet. Adm. tech. Bull.* 10-507, Nov. 16, 1953.
5. Lorr, M., Jenkins, R. L., & O'Connor, P. Factors descriptive of psychopathology and behavior of hospitalized patients. *J. abnorm. soc. Psychol.*, in press.
6. Moore, T. V. The essential psychoses and their functional syndromes. *Stud. Psychol. Psychiat. Cathol. Univer. Amer.*, 1933, **3**, No. 3.
7. U. S. Veterans Administration. Development of the Northport Record. *Infor. Bull.* 10-5-21, Psychiat. Neurol. Issue, July, 1950.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

Group Therapy Techniques with Hospitalized Chronic Schizophrenic Patients

Joseph M. Sacks and Stanley Berger
Franklin D. Roosevelt VA Hospital, Montrose, N.Y.[1, 2]

Until recent years, the predominant approach to group treatment of hospitalized schizophrenic patients called for mass activities or didactic lectures [4, 5, 7, 8, 11]. More recently, there has been an increased application of psychodynamic principles and techniques in group psychotherapy with psychotics [1, 3, 6]. Most dynamic approaches to group psychotherapy have in common the expression and release of feelings, mutual understanding and emotional support, recognition of the individual's highly personalized needs, an opportunity for each member of the group to test his behavior, thinking, and feeling against the reality of his impact upon others, a sharing of common problems with others, and the cooperative working out of solutions to these problems with ultimate self-understanding and acceptance. The authors of this paper believe that the group treatment of psychiatric patients involves all of these elements and something in addition which is of crucial importance: the understanding and exploration of distortions in the perceptions and behavior of the patients in their interpersonal relationships. This paper presents a description of the techniques employed and results obtained in an attempt to apply psychodynamic principles to the group treatment of chronic schizophrenics.

The techniques to be described have been applied for the past year to four groups of 28 patients on the Continued Treatment Service of the Franklin D. Roosevelt VA Hospital. The groups were conducted by cotherapists, including a psychiatrist, a staff psychologist, and a clinical psychologist trainee. The groups met twice weekly for periods of an hour and a half.

A control group was derived from the hospital admissions records of patients admitted to the Continued Treatment Service on the same day as the patients of the therapy group. No significant differences were found between the two groups with respect to age, socioeconomic level, education, or intelligence.

Theoretical Considerations

In our approach to group therapy, we incorporated some of the more basic theoretical formulations that have been fruitful in treating less severely disturbed patients. In general, the concepts discussed above were integrated into a philosophy of therapy which we felt would be of value in any group psychotherapy. However, the focusing of treatment upon chronic, regressed schizophrenic patients necessitated a broadening and elaboration of some of the more succinctly presented theoretical concepts, as well as additional formulations peculiar to our patient group.

We have been guided by the following assumptions:

1. One of the primary characteristics of schizophrenic patients, according to certain workers [9, 10, 12], is their acute sense of deprivation; that is, they react as though they have experienced the trauma of never having had, or of losing, a mother-figure who normally would have satisfied their physical needs (often symbolized by food), as well as their emotional needs, by providing comfort and consistent love, protection, and security. In reaction to these basic deprivations with their inherent threat to life itself, the infant has been forced

---

[1] Submitted with the approval of the Chief Medical Officer, Veterans Administration, who assumes no responsibility for the opinions expressed here, which are those of the authors.

[2] The authors gratefully acknowledge the participation and supervision of the following ward psychiatrists: Leopold Jordani, Maurice Pachter, and Paul Wenger. We also wish to thank Edwin Barker, Hanna Husserl, Warren Fox, Nathaniel Wagner, and Helen Singer, clinical psychologist trainees, who participated in various phases of this project.

to draw upon modes of adjustment and perception which are highly distorted and which, persisting to adulthood, prevent the development of more healthy and satisfying psychological mechanisms of defense.

2. The reliance upon inappropriate and distorted defenses is experienced by the patient as highly inadequate, fraught with potential danger and anguish, and in constant threat of collapse. Federn [2] stresses the point that the schizophrenic patient is constantly defending himself against becoming aware of the original fear of isolation and deprivation (death), and must continually apply whatever defenses he can to manage some sort of psychic equilibrium or reduction of tension. However, the patient is aware (at some level of consciousness) that these defenses are not effective.

3. Many techniques applicable to neurotic groups, or even to groups of ambulatory schizophrenics or schizophrenics in remission, are not effective in treating chronic schizophrenic patients. Collectively, these patients lack sufficient integration or group cohesion so that techniques making use of appeals to reason or judgment ordinarily fail. In the same manner, attempts to utilize the patients' ability to test reality or to control their impulses rarely meet with therapeutic benefit.

4. The most immediate task in group therapy with chronic schizophrenics is to build bridges of communication among members of the group and between patients and therapists. The inability to establish and maintain healthy and satisfying interpersonal relationships makes the schizophrenic patient extremely difficult to treat. Techniques that permit the therapist to reach the patient should be used and nurtured in the belief that they provide healthy avenues to the personality; this gives the therapist a valuable, though delicate, hold upon the inner desperate hope for health which, it is assumed, almost every patient has.

We believe that the techniques described below are valuable in a chronic-schizophrenic group setting, and that their application provides a successful means of treating these patients. In the following section we shall enumerate briefly the specific techniques and the theoretical rationale that have been employed in connection with various important components of the therapeutic situation. When possible, specific examples drawn from tape-recorded transcriptions of group sessions are included.

Technique

The therapy. Making contact with the patients requires a very active role on the part of the therapist, for contact cannot be accomplished by a passive-permissive attitude no matter how kindly or well intentioned. Silence, for example, begets more silence. We have seen a therapy group sitting together for an hour without uttering a single word. These patients, often withdrawn into their own autistic fantasies and reveries, cannot be brought into direct contact with each other without active stimulation, most frequently by a direct discussion of the silence. The following transcription demonstrates the active role of the therapist in building up a positive relationship in a patient who has exhibited marked hostility and assaultiveness:

P. You make it sound as if my father didn't give a damn about me.

T. That's how you feel?

P. My grandfather was someone I could get along with much better. He was the kind of fellow who gave advice and respected my opinions. He's a man who would shake your hand.

T. I will shake your hand.

P. You would?

T. Yes.

P. You mean as a sign of respect?

T. Yes.

P. Well then, here.

Patient extends hand; therapist walks over to patient and shakes his hand. Patient bursts into tears.

Another fruitful technique for active stimulation is to facilitate the patient's understanding of the deprivations which he experienced early in life and the manifestations of such needs in the present situation. Material which has meaning on other levels of adjustment is most frequently dealt with on this early level, assuming that this is most fruitful and meaningful for these patients [9]. If less direct methods are employed, patients respond with evasive, confused, autistic material, often engaging in long meaningless monologues. When the patients are able to resolve some of the difficulties in their earliest relationships, increasingly greater use is made of material related to later periods of adjustment. However, the early deprivations have been so severe that there is a constant return to problems arising out of this period. The handling of these feelings of infantile deprivation, in which the underlying fear of death is revealed, is demonstrated by the following example:

P. I would be all right if it weren't for the pain.

T. What pain?

P. It grips my throat, it hits me here [cheek] and my teeth lock together.

T. When does this happen?

P. It's from the whisper.

T. What whisper?

P. They say those nasty things. I have to fight them. I'd be all right if I didn't have to eat that medicine.

T. What medicine?

P. The stuff they put in my food. It makes my body jolt. I get these awful pains.

T. Why do you eat the medicine?

P. I have to. It keeps me alive. The doctors made my mother give it to me when I was a child so I wouldn't end up in a hospital.

T. You're in a hospital now.

P. Yeah, but I'm not dead.

T. You say you were afraid that you would die if you didn't eat the bitter medicine your mother gave you?

P. You're not kidding, Doc!

In order to validate our understanding of the patient's acute sense of oral deprivation in terms of the theoretical considerations discussed above, and our willingness to try to satisfy them, we bring "food" (chewing gum and chocolate) to the group sessions. It is not merely the direct physical satisfaction of oral cravings that is involved here, but more significantly, the process of receiving food from the parental figures as represented by the therapist, and partaking of it in the family setting of the group. The patients respond quite favorably, and it is noteworthy that in some cases the patients talk about buying candy for the therapist, or offer him cigarettes during the session. At one session a patient brought in a bun from his breakfast and placed it on the table in the center of the group for all to share.

In a further attempt to construct some relationship with the patients, the therapist may resort to overt attempts to dress in a similar fashion (no tie, sleeves rolled up), sit in a similar manner (slouched, or at times stiffly), and speak in a similar manner. In terms of the latter, the use of common language is certainly more meaningful to these patients, and at the same time is, in a sense, granting them permission to speak freely and honestly. The following example points out the importance of communication with the patient in his own language as a means of building up positive relationship:

P. I don't know how many personalities I have.

T. Why don't you know?

P. Because I always thought a person could talk to another person in an understandable way. That is personality.

T. Can you talk to a person in an understandable way?

P. I mean if he is angry, or if he is sad. Personality to me seems like I feel when I have medicine; the way you talk to a person, the way you answer.

T. It must have something to do with the way you talk to me. How do you talk to me?

P. I mean you talk not only because your words are well spoken, or the sound of your voice, but you can understand what I mean, and I can understand what you mean.

T. We understand each other.

P. Sort of a love affair. That's what I think personality is.

T. You love me because I understand you?

P. That's right.

The therapist. A difficult task for the therapist is to permit a relaxed and receptive atmosphere within himself so that he can be more sensitive to the highly symbolic associations of the patients, and can rely upon his own clinically intuitive responses to aid in communicating with them. To maintain this sensitivity to communication requires a healthy personality for the therapist, who can, by accepting his own impulses and needs in a mature manner, be more accepting of the patient's desires and impulses. It has been often observed that schizophrenic patients are very perceptive, and they will be quick to observe in the therapist anxiety and fear, and especially rejection.

The therapist maintains himself as the central figure in the group about which therapy pivots. Though the patients may shift about, the therapist always sits in the same seat, often indicating the strength and valence of the relationships, the patient's need for contact, and his desire for communication.

The consistency of the therapist is further emphasized by his remaining with the group despite other hospital assignments or commitments. If the group is transferred to another building, it is desirable for the therapist to be transferred likewise. Such maintenance of group integrity provides a group-stabilizing force, as well as a means of helping the patients make these kinds of moves and new adjustments.

The therapist continually attempts to build up positive feelings and tries to deal with the negative ones. Hostility, rivalry, and competition within the group must not be allowed to continue too long. When too much hostility is expressed by one patient to another it must be redirected (preferably to the therapist) and reduced in strength so that the angry patient and the rest of the group are not overwhelmed by an overt display. Prolonged uncontrolled expression of strong feeling is potentially upsetting to schizophrenic patients who are in constant dread of losing control over themselves, and it is possible for one agitated patient to set off the entire group if the therapist fails to be the strong, controlling figure. The following transcription illustrates bringing patient to patient hostility out in the open and redirecting it toward the therapist:

A. [For about fifteen minutes has been playing a didactic therapist role toward another patient, B, newly admitted to the group, asking questions about his history, his plans for the future, and finally giving a case summary.] So it looks as though you want to depend on Uncle Sam for the rest of your life. Don't you see that this dependency is part of your sickness?

C. [Has been squirming in his seat, flushing, left the room and returned.]

T. [to C] How have you been feeling these last few minutes?

C. Like I'm going to explode.

T. Why is that?

C. Listening to A shoot off his mouth. He isn't trying to help B, he is just trying to show how much he knows.

A. That's not true. We are supposed to help each other here, aren't we doctor?

T. Maybe you're doing to B what you feel I have done to you in the past, pointing out your own dependency needs. You must have felt pretty angry when I did that.

B. May I say something? Yes, I did resent what A was doing but also I felt he was trying to help me and I appreciate it.

C. He wasn't trying to help. He was just trying to play the doctor.

A. Maybe you've got something there, C. I certainly wasn't aware of it though. I guess you do resent it, doctor, when someone points out your own weaknesses to you.

[After the session A and C go out together talking with much animation, but in friendly fashion rather than in a hostile manner.]

The group. The patient's illness is thought of in terms of "ego-sickness" [2]. When patients request an answer to some realistic question, or when it is felt beneficial to discuss some aspect of the schizophrenic process, the explanation is centered about the patient's inability to reconcile bodily impulses with an unrelenting superego, and the resulting disintegration or fragmentation of the personality structure. When the therapist feels that it is beneficial to reply to a question by a patient concerning his illness, the reply should be honest and straightforward. Such frankness quickly earns the patient's respect. Dealing with the patient's illness openly and honestly is reassuring, not threatening as illustrated below:

P. So my father told me to come up here for a couple of weeks to rest and to gain some weight. Well, I've had enough rest, I've been in this hospital for more than a year. But they don't want me to go home yet. I don't understand it.

T. The truth is that you have an emotional disturbance. That's why you are in this hospital. This is no rest home.

P. Oh, I guess my father knew what he was doing.

T. What do you mean?

P. I couldn't control my emotions when I was home. I was blowing my top all the time, couldn't keep a job, afraid of people. I had to come up here to straighten out all that.

T. What about not understanding why you can't get out yet?

P. I guess that will have to wait until I straighten myself out a lot more.

The administrative problems of ward and hospital, as they affect the patient, cannot in themselves be considered a part of group therapy. It should be quickly pointed out to the patients that the purpose of the therapy is not to grant passes or leaves of absence, assign patients to work details, or engage in other administrative duties of this nature. These factors are certainly considered by the therapists in terms of what arises in therapy. However, the therapist reinforces his position in the group as a person sincerely interested in helping the patients get better, rather than one who essentially grants or denies privileges.

The solidarity of the group is emphasized by assigning the patients to the same sleeping and eating quarters, the same work details, and in general keeping them together on many hospital activities. While these procedures may involve considerable administrative difficulty, they have been found to be therapeutically beneficial.

The making of tape recordings and playing them at the subsequent session has been found to be a useful technique for enabling the group to gain a more objective and less distorted view of themselves. A frequent reaction is surprise at the amount of feeling displayed about a given subject. Schizophrenic patients in particular seem to derive a great deal of satisfaction from hearing their own voices and are eager to listen. They will interrupt the playing of the recording to elaborate something which they have said, or to speak of other aspects of the discussion. At the end of playing one such recording, dominated largely by two patients, one of these two turned to the other and said, "Let's you and me take a back seat next time and give the other fellows a chance." The following shows how the review of a previous session by tape recording enables a patient to interpret his own symbolic utterance and to present new material for further therapeutic exploration.

> P. It felt like I was choking, like there was a rope around my neck. Maybe one of my ancestors was a pirate, no, maybe I was a pirate in a previous life and I was hung by the neck from the yardarm of a ship.
>
> T. Why were you being hung?
>
> P. I don't know. I never committed any crime. [Patient interrupts recording at this point.] The only crimes I have ever committed were against myself. I was thinking about how I threw my life away. That choking feeling has nothing to do with pirates. It was my own disgust with myself and the thought of suicide coming out in the form of a symptom. That's what it was.

Individual conferences with the therapist are available to the patients when they so desire or when the therapist feels that a patient might return to the ward in an agitated state. It was in one of these individual sessions that a patient confided, for the first time, that he had suicidal ideas. Others have been able to talk about fantasies or impulses which they had previously suppressed.

Results

In our initial meetings with these patients, on the more inactive wards in the hospital, many of them were very withdrawn, apathetic, autistic, uninterested in the environment, and unaware of their own lack of personal tidiness and crude behavior. In no case did the patients appear to have any understanding of, or any interest in, group therapy. The first few sessions in the group found them to be uncommunicative and withdrawn, they showed no interest in the other patients or the therapist, and mainly sat for the entire time lost in their own preoccupations.

At the present time we see most significant changes in the patients' interpersonal experiences and behavior. The group meetings are active and alive, often with several patients trying to talk at the same time. They are interested in themselves and their environment, and even more noteworthy, they have demonstrated an increasing interest in each other and in the therapist. In addition, many of the patients have become interested and engaged in hospital activities and work details. In one case, an entire group was transferred to a rather active ward on which they are all assigned to work details, engage in athletics, and regularly attend hospital parties and movies.

These patients have made definite advances in achieving better emotional adjustment, and they are highly motivated for and better capable of receiving intensive psychotherapy. We feel that the therapeutic techniques employed have enabled us to penetrate long-standing psychic barriers and to initiate more healthy and gratifying modes of adjustment in pa-

Table 1

Hospital Status of Psychotherapy Group and Control Group After One Year

Status	Control group	Therapy group	D_σ	p
On More Disturbed Ward	4 (14%)	6 (21%)	.80	.42
Unchanged	18 (64%)	5 (18%)	3.53	.0004
On Improved Ward*	1 (4%)	13 (46%)	3.39	.0008
Discharged or Trial Visit	5 (18%)	4 (14%)	.37	.70
Total	28	28		

* Wards requiring better adjustment to other patients and personnel, and more active participation in the hospital program.

tients who might have otherwise remained lifelong, regressed schizophrenics.

It is apparent from Table 1 that there are no significant differences between the therapy and control groups in the "More Disturbed" and "Discharged or Trial Visit" categories. On the other hand, in the "Unchanged" and "Improved Ward" categories, the differences in favor of the therapy group are very large. In the control group, 64% are unchanged, compared with 18% of the therapy group. This difference is significant at the .0004 level. In the therapy group, 64% are improved, compared with 4% of the control group. This difference is significant at the .0008 level.

When the status categories were combined into Improved (On Improved Ward plus Discharged or TV) and Unimproved (More Disturbed Ward plus Unchanged), a chi square of 9.02 was found, indicating a probability of less than one in a hundred that the differences between the therapy and control groups are due to chance. These results are in the same direction as those obtained by Powdermaker and Frank [6, ch. XVII] and by Feifel and Schwartz [3].

Summary

This paper describes, in the opinion of the authors, some of the more effective group therapy techniques, which were actively applied for a period of one year to four groups of chronic schizophrenic patients in a large hospital setting. A clinical evaluation of the patients' improvement is offered, as well as statistical comparison with a control group.

Theoretical assumptions upon which these techniques have been based are presented. The techniques themselves are discussed in terms of the therapist, the group, and the therapy process.

Illustrations from recorded transcriptions of group therapy sessions are given as concrete illustrations of several of the points presented.

It is hoped that the presentation of this material will stimulate greater interest in the chronic schizophrenic patient, and that continued efforts will eventually discover therapeutic techniques that will prove fruitful and efficacious in restoring these patients to a more satisfactory way of life.

Received November 20, 1953.

References

1. Burchard, E. M. L., Michaels, J. J., & Kotkov, B. Criteria for the evaluation of group therapy. *Psychosom. Med.*, 1948, **10**, 257-274.
2. Federn, P. *Ego psychology and the psychoses.* New York: Basic Books, 1952.
3. Feifel, H., & Schwartz, A. D. Group psychotherapy with acutely disturbed psychotic patients. *J. consult. Psychol.*, 1953, **17**, 113-121.
4. Klapman, J. W. Group psychotherapy in institutions. *Group Psychother.*, 1951, **4**, 181-192.
5. Marsh, L. C. Group treatment of psychoses by psychological equivalent of the revival.*Ment. Hyg., N. Y.*, 1931, **15**, 328-349.
6. Powdermaker, Florence B., & Frank, J. D. *Group psychotherapy.* Cambridge, Mass.: Harvard Univer. Press, 1953.
7. Pratt, J. H. Results obtained in the treatment of pulmonary tuberculosis by the class method. *Brit. med. J.*, 1908, **2**, 1070-1079.
8. Pratt, J. H. The principles of class treatment and their application to various chronic diseases. *Hosp. soc. Serv.*, 1922, **6**, 401-411.
9. Rosen, J. *Direct analysis.* New York: Grune & Stratton, 1953.
10. Sechehaye, M. A. *Symbolic realization.* New York: International Universities Press, 1951.
11. Slavson, S. R. *Analytic group psychotherapy.* New York: Columbia Univer. Press, 1950.
12. Sullivan, H. S. *Conceptions of modern psychiatry.* Washington: W. A. White Psychiatric Foundation, 1947.

Journal of Consulting Psychology
Vol. 18, No. 4, 1954

The Influence of Fullness of Interview on the Reliability, Discriminability, and Validity of Interview Judgments[1]

Harry Levin
Harvard Graduate School of Education

Every interviewer has had the experience of talking both with laconic and with verbose respondents. For clinical purposes, the length of response may in itself be a significant factor. However, when interviews are used to gather research materials—that is, when the responses will be quantified and related to other variables—we may ask how the fullness of the responses affects the judgments made on the interview, assuming that both the short and the long replies are codable.

Factors determining the length of replies have not been widely examined. Both Shapiro and Eberhardt [4] and Feldman, Hyman, and Hart [1] found that there was wide variability among interviewers as to how much they probed the respondent. It is likely that the more extensive the probing, the longer the interview. Doubtless the general verbality of the respondent also influences the length.

Given interview responses of different lengths, or fullness, however, one may ask what influence this factor has on the reliability, discriminability, and validity of the coded judgments made from the material. A set of 379 transcribed interviews obtained in connection with a study of maternal child-rearing practices has offered an opportunity to make preliminary tests of three hypotheses.

Three kinds of codes were used to quantify the interviews: (*a*) item codes (henceforth labeled I) required judgments on responses to single specific questions; (*b*) summary codes (S) involved judgments based on a group of three or four questions; (*c*) over-all codes (O) took into account material which occurred not only in response to specific questions, but also at various points throughout the interview.

The following hypotheses were tested:

1. Reliability on all three types of codes is higher for the short than for the long interviews. The reasoning is that the added materials lead to different interpretations by the coders.

2. Discriminability of judgments is greater for the long than for the short interview. This prediction is based on the idea that more material in the interview permits finer discriminations among respondents.

3. Long interviews lead to more valid judgments than short interviews. It seems likely that added materials lead to judgments which are "truer" to behavior than are those based on sparser materials.

Procedure

Three hundred and seventy-nine mothers were interviewed about various aspects of their child-training procedures. A standardized interview schedule was used, consisting of a list of open questions which the interviewers were trained to use without change in wording and in the prescribed order. Following each of the main open questions was a series of suggested probes which the interviewer was free to use or omit, depending on the fullness of the answer to the original question. The interview took between one and a half to two hours, on the average.

The interviews were recorded on an Audograph and later transcribed into typescript.

---

[1] This article is based on data drawn from a study on "Identification in Young Children," conducted by the Harvard Laboratory of Human Development under the direction of Professor Robert R. Sears, and is supported by a grant from the National Institute of Mental Health (M-461). The author wishes to thank Dr. Eleanor E. Maccoby for her help with the article.

On the basis of the number of records used to complete the standard interview, two extreme groups of 50 respondents each were chosen. The "long" interview group was made up of 40 people who talked for approximately two and a quarter hours, and 10 who talked for more than two and a half hours. The "short" interview group contained 25 respondents who completed the interview in less than one hour, and 25 respondents who took about one and a quarter hours to respond to all of the questions. The two extreme groups chosen for this comparison differed by approximately one hour, on the average, in the length of the interviews.[2]

The 100 interviews considered in this report were taken by ten different interviewers. Two interviewers were responsible for 29, or 58%, of the short interviews, and one other interviewer took 18, or 36%, of the long interviews. It appears, therefore, that the length of the interviews may have been a function of two interdependent factors: the interviewer's technique and the amount that the respondent talked.

Typescripts were made from the records, and 188 judgments, or codings, were made from each interview. Two raters independently made the judgments on each interview and then reached a "pooled" judgment by discussing their ratings. Raters were rotated so that each of a group of six judges worked with every other judge an equal number of times. For the various analyses discussed in the present study, five scales each of the I, S, and O types were selected.

Results

The comparison of short and long interviews is based on the assumption that even the short interviews contained sufficient information for the raters to make a judgment. Out of the total of 750 pooled judgments being considered for each group, only 14 judgments in each group were not codable.

Reliability of judgments. Since two coders made the judgments on each interview, the correlations between their ratings prior to pooling provide an index to the reliability of coding. In Table 1 it can be seen that the reliabilities on the item scales were in every instance higher for the short than for the long interviews. On two of the item scales the differences between the reliabilities are statistically significant, and on two other scales the differences approach significance.

Table 1

Two-Coder Reliability of Item Codes Based on Long and Short Interviews

	Scales				
	I-1	I-2	I-3	I-4	I-5
Long interviews	.50	.65	.65	.52	.50
Short interviews	.86	.82	.74	.76	.72
$t_{r_1 r_2}$	3.57**	1.83	.84	2.01*	1.72

** Significant at the 1% level.
* Significant at the 5% level.

From Tables 2 and 3 it can be seen that, with one exception, the reliabilities of judgments on summary and over-all codes are similar regardless of the length of response. Scale O-4 dealt with the "mother's esteem of father" and was one of the most highly inferential scales used. It is interesting to note that when the basis for the judgment was scattered throughout the interview, was not based on any specific questions, and involved a great deal of inference by the coder, long interviews, providing more material, permitted higher agreement among the judges. In general, though, when the code is based on a specific response, emendations and elaborations of the response detract from intercoder agreement. On codes encompassing more material in the interview, length of response has little influence on reliability.

Table 2

Two-Coder Reliability of Summary Codes Based on Long and Short Interviews

	Scales				
	S-1	S-2	S-3	S-4	S-5
Long interviews	.74	.60	.44	.77	.83
Short interviews	.66	.59	.59	.75	.74
$t_{r_1 r_2}$	.75	.07	.99	.23	1.14

Discriminability of judgments. In research interviews, on the basis of which coded variables will be related to other variables, it is a prime requisite that the judgments discriminate maximally among respondents. With the variability of judgments as a criterion, the variances of the 15 codes on the long and short interview groups were computed, and the significance of difference between the "long" and "short" variances was computed. The results are given in Table 4. Of the 15 comparisons, three of the variances based on the long interviews are larger at the 5% level or better, and on one scale the variability of scores based on short interviews is significantly greater. However, the mean of the long interview variances is not significantly different from the mean variance of the short interview judgments (Festinger's $d = .86$) [2].

---

[2] Since the recorder disc revolved at a constant rate, it was possible for a faster talking respondent to say more in a given length of time. Actually, the correlation between length of interview as determined by number of discs used and the number of pages of the typescript is .75. The choice of extreme groups should minimize this source of error.

Table 3
Two-Coder Reliability of Over-all Codes Based on Long and Short Interviews

	Scales				
	O-1	O-2	O-3	O-4	O-5
Long interviews	.74	.58	.71	.72	.33
Short interviews	.82	.75	.75	.37	.42
$t_{r_1 r_2}$	.99	1.49	.41	2.49**	.50

** Significant at the 1% level.

Table 4
Variances of Judgments on Long and Short Interviews for Item (I), Summary (S), and Over-all (O) Codes

Scale	Variances long interviews	Variances short interviews	*F*
I-1	.89	1.14	1.28
I-2	1.41	1.58	1.12
I-3	1.01	.97	1.04
I-4	3.06	1.13	2.71**
I-5	1.00	1.16	1.16
S-1	.99	.94	1.05
S-2	2.40	2.53	1.05
S-3	.60	.70	1.17
S-4	1.33	.95	1.40*
S-5	2.25	1.78	1.26
O-1	1.70	1.56	1.09
O-2	.91	1.42	1.56*
O-3	2.15	2.20	1.02
O-4	1.87	1.44	1.30
O-5	7.82	5.07	1.54*

** Significant at the 1% level.
* Significant at the 5% level.

Validity of judgments. An ideal validity check would be to compare the judgments made on the interviews with the overt behaviors which the scales purport to measure. Since this was not feasible within the limits of the present study, a criterion of *logical* validity was used, that is, the utility of scores in generating predicted relationships with other variables. Based on the total 379 interviews, an analysis of social class differences in child-rearing practices was carried out.[3] It was found that 9 of the 15 scales being discussed in the current study were highly related to the social class memberships of the respondents. For example, there is more frequent use of physical punishment (Scale O-2) in the lower than in the middle class. Following the analysis of the total sample, specific predictions about the long and short interview groups were made as to how the variables would relate to social class membership. Within the limits of the reliabilities of the scales, it would be expected that the more valid the scales, the more clearly they will relate to social class.

To test these relationships, the long and short interviews were divided into higher and lower social class, and the means for each scale were calculated. In Table 5, the differences (d) between the means for the two social classes within each length-of-interview group are presented. The stronger the relationship to social class, the larger should be the d score. For example, for Scale I-4 (How Far is Child Expected to Go in School) the relationship is obviously higher for the short interviews, since there is a larger difference between the means, than in the long interviews (1.36 compared to .49).

If we sum the d scores separately for each type of scale, we find that there is a sizable d score for the item scales in favor of the short interview group; that is, the relationship of the variables to social class is higher for variables measured on the basis of the briefer interviews. These findings are in line with the higher reliabilities of the item scales, as reported in Table 1. It might be expected that, if the longer interviews permitted more valid judgments, the lower reliabilities of the long interview judgments might be somewhat overcome, thus equalizing the d scores. Such is obviously not the case.

---

[3] This study is reported in detail in [3].

Table 5

Means and Differences between Means on Selected Scales According to Length of Interview and Social Class of Respondent

Scale no.	Name	Short interviews			Long interviews		
		High social class ($N=25$)	Low social class ($N=25$)	d	High social class ($N=33$)	Low social class ($N=17$)	d
I-2	Permissiveness for masturbation	2.64	1.70	.94	2.74	1.91	.83
I-4	How far child expected to go in school	4.88	3.52	1.36	4.55	4.06	.49
I-5	Demonstrativeness: mother to child	3.38	3.32	.06 $\Sigma = 2.36$	3.32	3.24	.08 $\Sigma = 1.40$
S-1	Severity of toilet training	2.36	2.32	.04	2.53	3.24	.71
S-2	Mother's sex anxiety	3.02	4.66	1.64	3.36	4.56	1.20
S-4	Permissiveness for aggression toward parents	2.40	1.64	.76 $\Sigma = 2.44$	2.42	1.76	.66 $\Sigma = 2.57$
O-2	Use of physical punishment	3.16	3.90	.74	3.26	4.00	.74
O-4	Mother's evaluation of father	5.02	4.92	.10	5.17	4.26	.91
O-5	Agents of caretaking	3.26	2.20	1.06 $\Sigma = 1.90$	3.89	3.94	.05 $\Sigma = 1.70$

On neither the summary nor over-all scales did added length increase the accuracy of prediction. Actually, three d scores favor the short interviews; two, the long interviews; and on one scale the d scores are equal. The sums of d scores on the S and O scales are very similar. Therefore, contrary to our original expectations, on none of the three types of scales did added length of response appear to increase the accuracy of predictions.

Summary and Conclusions

Fifty long and fifty short interviews conducted with a standardized open interview schedule were selected, and various characteristics of the coder judgments made from the interviews were compared. The short interviews yielded as many codable responses as the long interviews. The following findings emerged:

1. When judgments are made on a specific section of the interview, intercoder agreement is higher on the short interviews. There are no differences in reliabilities between judgments on long and short interviews when the appropriate material is based on a group of questions scattered throughout the interview.

2. There are no differences between the ability of long and the ability of short interview judgments to discriminate among respondents.

3. When validity is judged by the utility of the ratings in generating relationships with other variables, long and short interviews are equally valid on lengthy and scattered materials in the interviews. There is a suggestion that judgments on a specific section of the interview are more valid in the short interviews, although this may be a reflection of the higher reliabilities of these judgments.

Received December 4, 1953.

References

1. Feldman, J. J., Hyman, H., & Hart, C. W. Interviewer effects on the quality of survey data. *Publ. Opin. Quart.*, 1951-52, **15**, 734-761.
2. Festinger, L. The significance of difference between means without reference to the frequency distribution function. *Psychometrika*, 1946, **11**, 97-105.
3. Maccoby, Eleanor, & Gibbs, P. Child rearing and social class. In C. B. Stendler & W. Martin (Eds.), *Readings in child development.* New York: Harcourt, Brace, 1954. Pp. 380-396.
4. Shapiro, J., & Eberhart, J. C. Interviewer differences in an intensive interview survey. *Int. J. Opin.-Attitude Res.*, 1947, **1**, 1-17.

NEW BOOKS AND TESTS

Books

Carmichael, Leonard. (Ed.) *Manual of child psychology.* (2nd Ed.) New York: Wiley, 1954. Pp. ix + 1295. $12.00.

The increments of knowledge from eight more years of research fully justify the publication of a new edition of the *Manual of Child Psychology.* The authors of the chapters differ considerably, however, in the extent to which they perceive forward movements in their fields. Some chapters are rewritten extensively with the citation of many new studies and the reappraisal of some older ones; others seem to have undergone only perfunctory changes since the 1946 edition. Among the most thoroughly revised sections are Munn's account of learning, McCarthy's of language development, Harold Jones's of the environment and development, Vernon Jones's of character, and Jersild's of emotion. Chapters almost unchanged are those by Carmichael on prenatal development, Cruikshank on animal infancy, and Pratt on the neonate. One wonders whether the statement in the preface that "development has been more active in certain fields" accounts for all of the differences. Three chapters are new, one of them an excellent account of adolescence by Horrocks. The other new chapters do not fulfill their potentialities. Benda, writing on the psychopathology of childhood, devotes 31 pages to a well-documented survey of mental deficiency, and then gives only 14 pages to a didactic and elementary account of psychoneuroses and childhood schizophrenia, citing hardly a research study. Harold and Gladys Anderson append over 200 references to their chapter on social development but use them mainly in lists of names and dates. The chapter is expository with a few illustrations from the authors' own research. The *Manual* as a whole, in spite of the inevitable unevenness resulting from multiple authorship, remains a useful guide to the research on childhood. Its direct value to the needs and concerns of clinical psychologists is limited. — *L. F. S.*

Gladwin, Thomas, & Sarason, Seymour B. *Truk: man in paradise.* New York: Wenner-Gren Foundation for Anthropological Research, 1953. Pp. 651 + 8 plates. $6.50.

Long aware of the values of anthropological studies for their work, clinicians will find this monograph particularly relevant and stimulating. In the first place, it represents at a still more sophisticated and insightful level the kind of collaboration between anthropologist and psychologist that made the Dubois-Oberholzer study of the Alorese so seminal for both disciplines. Second, since it is one of a series of investigations of Trukese culture, this report is unburdened with such specifically ethnological matters as linguistics, formal social structure, systems of property ownership, and the character of the material culture. Its content is concerned entirely with Trukese personality development and the sociocultural matrix within which it finds expression. Consequently, it deals with the problems of greatest moment to American clinical psychologists but from a point of view that permits a valuable perspective. Third, the book is crammed with extensive life histories and excerpts from projective test protocols. These materials will excite the imagination of all who use such data in their daily work and provoke fruitful hypotheses among the thoughtful for testing in both clinical and research contexts. This fat volume is at once lively and serious, and it contributes both to an understanding of the conditions affecting personality development and of the techniques by which it is conceptualized and investigated. — ***E. J. S.***

Greenacre, Phyllis. (Ed.) *Affective disorders.* New York: International Universities Press, 1953. Pp. 212. $3.00.

A symposium of five papers, mainly on depression, presented before the American Psychoanalytic Association in 1950 and 1951. Edward Bibring and Edith Jacobson are in agreement that depression is a helplessness reaction with loss of self-esteem, and both have difficulty in reconciling the view with some earlier analytic interpretations. Elizabeth Zetzel presents an extended apology for Melanie Klein's theories. George Gero gives an interesting case of anorexia as an equivalent of depression. M. Katan views both depressed and man-

---

Note.—The reviews were prepared by the Editor and Associate Editors, who may be identified by their initials.

ic symptoms as mechanisms of denial and restitution. — *L. F. S.*

Harms, Ernest. *Essentials of abnormal child psychology.* New York: Julian Press, 1953. Pp. xiii + 265. $5.00.

In this book, the author presents his views of the genetic development of the child, before discussing therapy and diagnosis. He considers the concept of infantile sexuality quite inadequate, and the parental relation to be beyond sexual experience. The relationship to the father is described under the rubric of *paternus* and to the mother of *materna,* which he considers to be inherited archetypical factors comparable to Jung's *animus* and *anima.* His discussion of case material and specific diagnostic categories follows his conviction that different techniques are applicable to different cases, and that patients are best helped by psychotherapists who share their own psychic structure. He believes the incidence of true childhood schizophrenia greatly overestimated. His discussion of the diagnosis and treatment of feeblemindedness, and of the use of child art as diagnostic means, will be of particular interest to psychologists. — *A. R.*

Pennington, L. A., & Berg, Irwin A. (Eds.) *An introduction to clinical psychology.* (2nd Ed.) New York: Ronald, 1954. Pp. vii + 709. $6.50.

This is an excellent textbook, planned to give students an overview of the whole field of clinical psychology. Its editors have emphasized general principles rather than the mastery of special techniques. Although there is a long list of contributing writers, their varying viewpoints do not distract from the basic philosophy but rather make the student aware of the complexity of the field of clinical psychology. The second edition is not a reprint but a carefully considered revision. The chapters that are retained have been revised, and at least ten new chapters are added. It should be valuable for an introductory course in clinical psychology. — *B. M. L.*

Pepinsky, Harold B., & Pepinsky, Pauline Nichols. *Counseling theory and practice.* New York: Ronald, 1954. Pp. viii + 307. $4.50.

The Pepinskys resolve a familiar dilemma by integration. Counselors can and must be scientists as well as practitioners; they do not have to choose between the roles. Furthermore, being a scientist is not limited to an isolated part of one's life labeled "research," but is an essential feature of the counseling process itself. Like an experiment, each client presents a challenge to observe, to infer, to relate to a general framework of theory, to formulate particular hypotheses, and to assess the outcome. The authors wisely refrain from preaching an acceptance of their analysis. Instead, they illustrate it persuasively. The place of formal theory is shown by a learning approach to client behavior which uses Hull's symbols and also draws heavily on Mowrer, Berlyne, Skinner, Dollard and Miller, and Brown and Farber. Observation, inference, hypothesis making, and assessment are illustrated by case studies of counseling. Assessment is approached realistically, with a cogent discussion of criterion variables. This book may help to make counselors—and psychotherapists—more aware of the assumptions that they bring to their contacts with clients. As a text, it may help students to see the intimate relationships between science and practice. — *L.F.S.*

Stott, D. H. *Saving children from delinquency.* New York: Philosophical Library, 1953. Pp. x + 266. $4.75.

Stott has placed the emphasis upon "saving" children from delinquency—by prevention if possible, and by avoiding further difficulty after delinquency has occurred. The illustrations and situations are typically British, but the American reader will find much to help him in meeting his own problems since the basic needs are similar. Visiting teachers, probation officers, and social workers will find the book very readable and stimulating. — *B. M. L.*

Vedder, Clyde B. *The juvenile offender: perspective and readings.* Garden City, N. Y.: Doubleday, 1954. Pp. xii + 510. $6.00.

As the subtitle indicates, this book is a collection of readings from periodicals dealing with the subject of juvenile delinquency. The 71 reprinted articles are organized into 13 chapters, under headings like "The Extent of Juvenile Delinquency," "Economic Conditions and Familial Factors," "Special Personality and Behavior Problems," "Probation," and the like. Each chapter starts with a brief discussion of the particular aspect of juvenile delinquency by the author of this volume, well documented from authoritative treatments, and followed by the periodical readings. Twelve social science periodicals are represented in the reprints of the 71 articles but 90% are from four journals, and 60% from one: *Federal Probation.* In reading the volume, the impression is gained that the entire book is a republication of the last-mentioned periodical. Nevertheless, the variety of articles is great, and most of the authors are outstanding authorities in the field. It is a valuable source book for the student of juvenile delinquency. — *M. K.*

Wolffheim, Nelly. *Psychology in the nursery school.* (Trans. by Charles L. Hannam.) New York: Philosophical Library, 1953. Pp. 144. $3.75.

Psychology in the Nursery School is a translation of a work originally published in German which pleads for a psychoanalytically oriented nursery school. Dr. Wolffheim feels that Freudian psychology has an important contribution to the education of young children, and should lead to new ways of studying them. She believes that psychoanalysis would benefit both the teacher and her educational work; in fact she feels it would be a desirable part of the professional training of teachers. The book is challenging but not very practical. The author admits that her point of view is based on theoretical studies and experiences in small nursery schools only. — *B. M. L.*

Tests

Burgemeister, Bessie B., Blum, Lucille Hollander, & Lorge, Irving. *Columbia Mental Maturity Scale.* Ages 3-12. Individual test. 1 form. Untimed, (15-30) min. Kit of 100 test cards, with manual, pp. 13 ($35.00); record blank (85¢ per 35). Yonkers, N. Y.: World Book Co., 1954.

A seven-year research project culminated in the development of this unique individual intelligence test designed for the needs of children with cerebral palsy and other severe motor or speech handicaps. The materials are 100 large cards, 6 by 19 in., each containing from three to five drawings, many in color. The task is the same for each item: to select the picture that does not belong with the others. Most children point to the figure; handicapped children who can only nod or make a sound can also be examined successfully. The test was standardized on 957 normal children from 3 through 12 years of age, and also tried with 147 cerebral palsied children. In single-year age groups of normal children, its corrected odd-even reliability is from .89 to .92, and its correlations with the Stanford-Binet range from .66 to .88, with an average value of about .75. The test will surely find wide use in clinical work and research with the severely handicapped. It is not likely to replace existing instruments for testing normal children. The very features of simplicity that make it suitable for the handicapped hinder the useful auxiliary observations often provided by verbal tests and performance tests with varied tasks. Even so, another well-standardized intelligence test is a real asset to clinical psychology, and may find uses beyond its authors' original intentions. — *L. F. S.*

Eisenson, John. *Examining for aphasia.* (Rev. Ed.) Adolescent-adult. Individual test. 1 form. Manual with plates, pp. v + 73 ($3.50); record form, pp. 12, ($3.50 per 25); set ($6.00). New York: Psychological Corp., 1954.

Eisenson's revised tests for aphasia do not differ greatly from their first edition of 1946. The broad scheme of testing remains the same, with some improvements in the pictorial materials, in the arrangement of plates, and in the selection of paragraphs for oral and silent reading. An examination for aphasia is necessarily qualitative, and the tests therefore cannot be judged by the usual standards for reliability and validity. Some approximate quantifications of these tests would be helpful, however, at least to give an examiner clues for judging whether the disability in a given area is "moderate" or "severe." The manual is more extensively revised than the test itself, the first 28 pages comprising a miniature textbook on nature, symptoms, and varieties of aphasia, and the examination of aphasic children. Perhaps because of its condensation, this introductory material tends to be arbitrary and uncritical. Still, it provides a useful check list of definitions and procedures for the use of experienced psychologists. — *L. F. S.*

Wesman, Alexander G., *Verbal Test,* & Doppelt, Jerome E., *Numerical Test. Personnel Test for Industry (PTI).* Trade school-adult. 2 forms, each test. 5 (10) min., verbal; 20 (25) min., numerical. Test booklet, each test ($1.60 per 25), with manual, pp. 10, and keys; specimen set (35¢). New York: Psychological Corp., 1952, 1954.

The verbal and numerical tests introduce a new series of *Personnel Tests for Industry* which is to include tests of other abilities and skills. The verbal test is a 50-item, multiple-choice vocabulary test, and the numerical test a series of 30 arithmetic problems based on industrial tasks. The manual, modestly designated as preliminary, gives extensive information about reliabilities and the prediction of criteria, based on industrial workers, and percentile norms derived from several industrial samples and one trade school sample, totaling some 1400 cases. Although conventionally conceived, the tests show the application of high professional standards. They illustrate the trend toward the construction of tests especially designed for adults in industry. — *L. F. S.*

Books Received

Abramson, Harold A. (Ed.) *Problems of consciousness.* New York: Josiah Macy, Jr. Foundation, 1954. Pp. 177. $3.25.

Anastasi, Anne. *Psychological testing.* New York: Macmillan, 1954. Pp. xiii + 682. $6.75.

Burmeister, Eva. *Roofs for the family.* New York: Columbia Univer. Press, 1954. Pp. 203. $3.25.

Caplow, Theodore. *The sociology of work.* Minneapolis: Univer. of Minnesota Press, 1954. Pp. viii + 330. $5.00.

Commins, W. D., & Fagin, Barry. *Principles of educational psychology.* New York: Ronald, 1954. Pp. xvi + 795. $5.75.

De Sade. *Selected writings.* (Trans. by Leonard de Saint-Yves.) London: Peter Owen, and New York: British Book Centre, 1954. Pp. 306. $6.75.

Freud, Sigmund. *The origins of psychoanalysis: letters to Wilhelm Fliess, drafts and notes. 1887-1902.* New York: Basic Books, 1954. Pp. xi + 486. $6.75.

Fryer, Douglas H., Henry, Edwin R., & Sparks, Charles P. *General psychology.* (4th Ed.) College Outline Series. New York: Barnes & Noble, 1954. Pp. xix + 300. $1.50 (paper).

Gorer, Geoffrey. *The life and ideas of the Marquis de Sade.* London: Peter Owen, and New York: British Book Centre, 1954. Pp. 244. $3.50.

Martin, John Bartlow. *Break down the walls.* New York: Ballantine, 1954. Pp. 310. $3.50 (cloth), 50¢ (paper).

Pitt, William J., & Goldberg, Jacob A. *Psychology.* New York: McGraw-Hill, 1954. Pp. 414. $4.50.

Remmers, H. H., Ryden, Einar R., & Morgan, Clellen L. *Introduction to educational psychology.* New York: Harper, 1954. Pp. ix + 435. $4.00.

Runes, Dagobert D. *Letters to my daughter.* New York: Philosophical Library, 1954. Pp. 131. $2.50.

Ryan, Thomas Arthur, & Smith, Patricia Cain. *Principles of industrial psychology.* New York: Ronald, 1954. Pp. xiv + 534. $5.50.

Schwing, Gertrud. *A way to the soul of the mentally ill.* New York: International Universities Press, 1954. Pp. 158. $3.00.

Wulfeck, Joseph W., & Bennett, Edward M. *The language of dynamic psychology as related to motivation research.* New York: McGraw-Hill, 1954. Pp. 111. $4.00.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

An Assessment of the Current Role of Psychodiagnostic Testing[1]

Alan K. Rosenwald
University of Illinois, College of Medicine

In any attempt to consider the current role of psychodiagnostic testing in clinical psychology, one can see analogues in the divergent opinions and shades of thought characteristic of our political scene.

Our various media of current opinion have given much, perhaps too much, consideration to the varied political philosophies adopted by this or that person. Liberal, conservative, nationalist, internationalist, leftist, rightist, etc. have become common terms of approval or opprobrium. To a lesser degree, it seems that psychologists are just as likely to stereotype their fellow psychologists by some equally superficial characteristic. The clinician decries the experimentalist as lacking in certain qualities, chiefly human; the experimentalist condemns the clinician by use of the telling, if overly used, phrase, "He's no scientist." Clinicians war with other clinicians so that "the battle lines are drawn." Projectivists clashing with the non-projectivists, objectivists with the intuitionists, statisticians with the qualitative analysts, and Freudians with the non-Freudians are illustrative of the strife in which each and everyone of us becomes involved.

In this evaluation of the current major trends in diagnostic testing and in trying to anticipate our hopes or expectations of the future, the writer, too, finds himself forced into the use of labels. With this injected note of warning about possible oversimplification, the writer would wish to consider the trend, or label, illustrated by the school of thought he has designated as conservative. The conservative seeks to demonstrate the validity of his diagnostic instruments. In fact, the concern seems more than primary; it seems to be an all-encompassing one. This evaluative statement should not be construed as an attempt to decry validity and/or reliability studies, nor is it a passing glance at superficially recognizing the need for such studies only to embark on a series of psychological speculations which lack any scientific basing. It is a matter of concern, however, that the literature abounds in numerous studies which by their design seem almost destined to purposefully devalue certain tests.

As an example of the validational or so-called scientific approach, a study by Wittenborn illustrates the grievous principle when he says, "One of the outstanding characteristics of the Rorschach responses is their highly individual quality, their extreme variability from person to person. This makes conventional statistical treatment of Rorschach data awkward because of the difficulty in securing a large number of roughly identical responses. For example, out of a group of one hundred normal subjects many, if not all, would report human movement responses but relatively few would have more than two, or at the most, three identical responses in common. In the present study an attempt has been made to reduce this difficulty by the use of the Harrower-Erickson checklist" [6, p. 258]. Wittenborn then notes that some question has been raised as to the validity of this check list for diagnostic purposes, but defends his use in this particular study since he wants to investigate the interrelationship of various scoring factors, specifically, movement and color. It would seem that such a defense is, at best, questionable, since if the check list is invalid for diagnostic purposes, then one would logically question whether or not the responses gained from such a procedure could permit one to make sound inferences about the Rorschach method in its clinical setting. Moreover, it would seem that there is a potential danger to the reasoning which Wittenborn uses. According to his reasoning, if we have a particular test which fails to lend itself to being treated by orthodox statistical procedures, then the test itself must be altered so as to render it suitable for statistical manipulation. The reasoning is something like this: a switch to an invalid check list permits the data to be processed statistically. In other words, if a particular phenomenon can't be measured, measure an artifact of the phenomenon. On the basis of data gained from measuring the artifact, one can then turn to generalizing about the original phenomenon.

---

[1] Presented as part of the symposium, "Assessment of Current Trends in Psychodiagnosis," at the APA meeting, Cleveland, September, 1953.

As illustrative of the dynamicist or the "radical" approach, one can turn to Gurvitz's volume, *The Dynamics of Psychological Testing* [3]. In a particular instance, he is evaluating the association a patient made to a female figure drawing. When asked for the worst part of the drawing, the patient replied "Toe nails; they're ingrown, that's why she's not wearing shoes." Later on, in his analysis of the test protocol, Gurvitz says, "The ingrown toe nails appear to represent the self-punishing and masochistic elements in the personality" [3, pp. 116-119]. The existence of masochistic and self-punishing elements in a personality as stated is of no moment. Actually it would be a far more unusual statement in a psychological report if one could read that patients' masochistic and self-punitive elements were of little import in their personality. Certainly it is difficult to rule out the possible existence of such elements in personality formation; one can, however, question the basis on which the inference was made. Before reading his conclusion, the writer would have assumed that the patient was possibly attempting to be facetious and, by the same token, critical of the question or questioner. If speculation were the order, one might even say that the patient was trying to rationalize her economic deprivation, i.e., lack of shoes, by focusing on the somatic. At any rate, the elemental importance of the statement doesn't seem so great that it permits the delineation of a major aspect of personality.

Apparently many diagnosticians are doing what some psychoanalysts did a number of years ago (some still do) in attaching undue importance to everything that a person may say. Also, it would seem that we are in danger of covering our failure to understand certain personality operations by trying to couch our reports in high-sounding, esoteric language, which simply serves to make the person so described something less than a living being. One need not be unmindful of the danger that attempts to make reports too individualized may lead us on an unscientific road. If we get too enamored with our own literary creativeness so that we neglect to make certain general statements which should be made about an individual, then an injustice is being perpetrated. When these general statements become "he is an immature individual" or "his problems of hostility have not been solved" or what has now become the most cliché of psychological clichés, "his intellectual efficiency does not measure up to his intellectual potential," then it would appear that our reports have become jaded, lacking in specific statements about an individual, and are becoming a mass of generalities which unfortunately don't even glitter.

Holt's analysis of the TAT of Jay, a case study, illustrates too well the point. Before setting down to analyze the results, Holt gives his ESP interpretation. "As a kind of control, I decided to set down, before looking at the TAT some of the truisms and clichés that are found in most test reports, and that can be fairly safely said once you know someone is a patient of Jay's age and sex" [5, p. 457]. Then follows this ESP interpretation, which makes perfectly good reading and is just as definitive as many psychological reports that thrive in case files throughout the country.

Much of what has been said has concerned itself with the varying attitudes held toward projective tests. Many of the same observations can be made if we examine the area of intelligence testing. The following statements are found in Patterson's manual: "His social judgment is relatively superior, indicating that he has been capable of profiting by his experience. In contrast, however, foresight and planning ability (Picture Arrangement) is poor. This poor performance may have been affected somewhat by his difficulty of attention and concentration which may also have affected the Picture Completion and Digit Symbol as well as the Digit Span tests. Picture Arrangement and the Similarities as well as Vocabulary may also be affected by cultural background. In the Picture Arrangement test, after successfully arranging the first group of pictures from left to right, he arranged the next group from right to left and could give no explanation when asked to; he immediately placed the next series (the elevator) in a vertical order" [4, pp. 36-37]. Here again, we are confronted with interpretations that fail to account for discordant elements in the data as well as end up by denying the interpretation which has just been given. If we have an individual whose social judgment is relatively superior, then it would seem that his foresight and planning ability should also be superior. If, however, the performance of a patient is affected by both lack of concentration and attention, as well as poor cultural background, then perhaps we are not permitted to infer what an individual's foresight and planning ability are. Maybe we should say that he is so distractible that we could infer he could not plan ahead in a prudent manner. We must question whether we have a measure of a person's social judgment merely because he has a high score on Comprehension. Patterson does imply in the quotation above that his social judgment is an index to what the patient was like prior to his illness and that he had been capable of profiting from past experiences. If that is the case, why mention a poor cultural background, which would certainly connote a lack of experiences, common to all of us, from which he could profit. Further, why is there no emphasis on the confusion which surrounds the performance on the Picture Arrangement test and which would warrant considering it the production of a brain-damaged individual.

These examples from Patterson, Gurvitz, and Wittenborn have been selected at random from the literature to illustrate certain trends in diagnostic testing, rather than as three particular studies with which the writer wishes to quarrel. Certain conclusions should be drawn from these types of studies. First, it seems that our need for purification of our diagnostic instruments has led us all too frequently into neglecting how we get data, but then becoming extremely precise in our measurement of very unprecise primary data. Secondly, our need to demonstrate our "know-how" has caused us to engage in many speculative statements, of which a great number cannot be validated because we have no criteria by which to demonstrate their validity. In other words, their world-wide applicability to man makes our individual patient no different from anybody else in the world. Finally, we apparently seem unable to ever categorically and unequivocally state that we don't know the answer to a particular question — in other words, that we just don't have all the answers to all the questions. Rather we seem to get involved in a series of statements, some of which say one thing, followed by statements which deny what we have just said. Perhaps this is a tribute to our capacity for tolerating ambiguity; it does not, however, make for definitive diagnostic interpretations.

One should pause and consider some of the uses of diagnostic testing in a psychiatric setting. It would lead us far afield to consider the many different kinds of problems upon which we may be asked an opinion. We can say most referrals in a psychiatric clinic seek to know something about a person's intelligence, both with respect to over-all capacity as well as strengths and weaknesses shown in the separate functions, something of the personality structure, some of the attitudinal constellations which are dominant in a person, etc. More broadly than this, however, is the implied question, "What sort of a person is this?" Presuming the person has pathology, what sort of expectation do we hold regarding the course of illness; what can we expect in therapy? The writer feels that most referrents are not primarily interested in how we arrive at conclusions, providing, of course, that we do not damage the patient in the process of reaching our conclusions. They are interested in our conclusions, either because they constitute a check on their subjective impressions or because they reveal areas of personality functioning which they had not been able to explore adequately. There is danger of belaboring the obvious, but it seems well to consider what we are doing before we try to say what we should be doing. If we boil down the major functions of diagnostic testing as aiding in differential diagnosis, saying something about prognosis in or without therapy, and always recognizing that diagnostic instruments are useful research tools, why the need to assess the role of testing? The need to establish its role grows from a need to define its value. This concern is enhanced by the declining emphasis on diagnostic testing. In recent years clinical psychologists have shifted overwhelmingly in the direction of becoming therapists — and diagnostic testing has become confined by the "expert" telling the "nonexpert" how to examine and how to interpret results, what to do, what not to do, etc. Rarely does he do what he teaches. Moreover, the "average" person who comes for training in clinical psychology usually has in mind training in therapy or doing research, and only secondarily training in the diagnostic study of patients. The reasons which underlie the subordination of diagnostic testing to therapy and research are too numerous and involved to consider here. At any rate, the result is evident. Diagnostic testing may be likened to the status of a second-class citizen who is tolerated and endured but who is pegged below his superiors.

The writer questions strongly the desirability of this rejection of diagnostic testing for therapy. In turning to therapy, clinical psychologists are neglecting areas of study which could occupy their interest but which do not. This feeling does not stem from any antihumanitarian or antitherapeutic origin. Certainly psychological research in the therapeutic process will lead to some very valuable contributions. The criticism, however, is that there are problems in clinical diagnosis which require careful study, and that these will be neglected as the clinical psychologist focuses on being the practitioner.

Thus, all of us speak very seriously of the need to study the normal personality, yet our words are, for the most part, empty ones. Ask yourself how long it has been, if ever, since you wrote a report in which adequacy of function was the keynote of your report rather than a person's many inadequacies. We are now very proud of our newly acquired dynamic orientation which permits us to look askance at the work of Kraepelin, Bleuler, or any of the more descriptive psychiatrists. Instead, we speak knowingly of weak ego functioning, poor reality testing, latent schizophrenia, etc. until we are caught up in a series of ritualistic statements which, although colorful, fail to help us understand the patient.

We are victims of what Herbert Block, the *Washington Post* cartoonist, in his *The Herblock Book* calls *Mumbo-jumbo*. "One of the things that makes life complicated . . . is the practice of Mumbo-jumbo. Mumbo-jumbo, as I use the term, is the transference of devotion from a set of principles to the symbols and rituals that are built around them. It's the business of taking over a good sound idea, building an elaborate shell around it, then carefully removing the already wilting idea and beating its brains out with the shell. Mumbo-jumbo creeps up in practically every field of endeavor. It gets into the arts . . . into medicine when the operation becomes more important than the patients" [2, pp. 41-42].

Granted that we lose the patient in the process of studying him, why should this be? Partially, it is because we have failed to see the patient as he lives from day to day. In our assessment of people, we are seemingly at a loss in predicting such elementary things as how the person gets along with other persons, whether a patient is liked or disliked by other patients, what sorts of people he likes or dislikes, who the leader of a psychiatric ward might be, how the patient would get along with certain authority figures, etc., etc. It is not unreasonable to propose that validational studies should concern themselves with such problems. The writer concludes, a priori, that there is nothing wrong with psychologists spending time observing the behavior of people. As an example, one can see much merit in the approach of Barker and co-workers [1]. The time-worn story about a patient's remark to a psychiatrist still holds. "What do you know about a patient? You only see him once, the nurses and attendants are here only eight hours a day, but me, I'm here a full 24 hours a day. What do you want to know about so and so?"

Increased knowledge of the ins and outs of a patient's life should lead us to try and cast our studies in a more practical mold. Not only that, but it might lead us to use our diagnostic instruments as a means of assessing the person as he lives from day to day rather than describing him by means of a series of generalized statements which are impressive and high-sounding, but not very meaningful.

In closing, the writer would like to re-emphasize his feeling that diagnostic testing should further our understanding of personality structure and add to our limited knowledge of psychopathology. If we glance back at what has been written, much has been couched negatively. This negativism, as it were, has its justification in a feeling that we are not pressing toward the solution of major problems in clinical psychology which could be aided by diagnostic testing. Our literature is swollen with papers reporting the administration of good clinical programs, validation or invalidation of countless tests, and innumerable brochures and manuals on interpretation of this or that test. We seem to be avoiding any long-term problems in clinical psychology; administrative planning and the mushrooming development of new tests with all the standardization problems lend our work a certain frenetic quality. Is it an idle hope that we will move in the direction of long-term studies which seek to answer certain basic problems.

Received February 22, 1954.

References

1. Barker, R. G., & Wright, H. F. *One boy's day.* New York: Harper, 1951.
2. Block, H. *The Herblock book.* Boston: Beacon, 1952.
3. Gurvitz, M. S. *The dynamics of psychological testing.* New York: Grune & Stratton, 1951.
4. Patterson, C. H. *The Wechsler-Bellevue scales, a guide for counselors.* Springfield, Ill.: Charles C. Thomas, 1953.
5. Shneidman, E. S., Bell, J. E., Fine, R., Holt, R. R., Kutash, S. B., & Vorhaus, Pauline G. The case of Jay: interpretations and discussion. *J. proj. Tech.*, 1952, **16**, 444-475.
6. Wittenborn, J. R. Statistical tests of certain Rorschach assumptions: analyses of discrete responses. *J. consult. Psychol.*, 1949, **13**, 257-267.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Acceptance and Acceptability: An Examination of Social Reciprocity[1, 2]

Seymour L. Zelen
State University of Iowa

One hundred and forty-five sixth-grade children in five classes were given the Bonney Sociometric, the Feelings of Personal Worth Subscale of the California Test of Personality, and the W-A-Y technique. The latter two instruments yielded scores of self-acceptance, while the sociometric gave scores of peer status and acceptance of peers.

Peer acceptance correlated .59 with acceptance of others, significant at the .01 level, and .30 and .39 with the self-acceptance scores of the W-A-Y and the CTP Feelings of Personal Worth Scale respectively, both significant at the .01 level. Acceptance of others correlated .08 with the W-A-Y and .10 with the Personal Worth Scale; neither of these proved significant.

The expected positive correlations of peer acceptance with acceptance of others and acceptance of self were obtained. This suggests that liking others and being liked by others are reciprocal characteristics. Individuals with high group status tend to find many group members highly acceptable. Self-perception and perception by one's peers also produced a significant relationship. These findings suggest that a child who has positive feelings about himself is better able to devote his energies to the group activities and to cooperate more fully with others. The child with negative self-percepts may be constantly on guard against new perceived threats from others.

The lack of any relationship between the measures of self-acceptance and acceptance of others was unexpected. An interpretation is offered in terms of the discriminability of interpersonal or social cues.

It is possible that children might be aware of and capable of discriminating cues of external behavior, but would not be aware of the cues upon which the more subtle social relationships are based.

Brief Report
Received June 24, 1954.

References

1. Bonney, M. E. Personality traits of socially successful and socially unsuccessful children. *J. educ. Psychol.*, 1943, **34**, 449-472.
2. Bugental, J. F. T., & Zelen, S. L. Investigations into the 'self-concept'. I. The W-A-Y technique. *J. Pers.*, 1950, **18**, 483-493.
3. Clark, W., Tiegs, E., & Thorpe, L. California test of personality: Form AA, Intermediate. Los Angeles: California Test Bureau, 1953.
4. Sheerer, E. An analysis of the relationship between acceptance of and respect for the self and acceptance of and respect for others in ten counseling cases. *J. consult. Psychol.*, 1949, **13**, 169-175.

---

[1] An extended report of this study may be obtained without charge from Seymour L. Zelen, Child Welfare Research Station, University of Iowa, Iowa City, Iowa, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4307 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[2] This research was supported by Research Grant MH-301 from the National Institute of Mental Health, of the National Institutes of Health, U. S. Public Health Service.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Rorschach Color Responses and Manifest Anxiety in College Women

Wayne H. Holtzman, Ira Iscoe
The University of Texas
and A. D. Calvin[1]
Michigan State College

In recent years there has been a sharp increase in the number of experimental studies concerning the validity of the Rorschach test. In particular, a group of investigations has appeared dealing with the nature of reactions to color. In reviewing these studies, Benton [2] and Hertz [9] point out the generally negative findings with respect to the effects of color and emphasize the need for considerable revision of earlier Rorschach theory. Most of these studies are limited in that they dealt with the validation of only one or two Rorschach scores or considered only one of many pertinent independent variables.

The present research was designed to investigate the effects of anxiety, color, stimulus configuration, order of presentation, and examiner, considered singly and in combination, upon a large number of Rorschach variables.

Description of the Study

The Minnesota Multiphasic Personality Inventory (MMPI) was administered to 696 college women in various dormitories and sororities at The University of Texas. Answer sheets were scored for both the Taylor Manifest Anxiety Scale [17] and the Winne Scale of Neuroticism [20]. High- and low-anxiety groups were established by selecting women having both Taylor scores in the upper or lower 15 per cent and Winne scores in the upper or lower 25 per cent of the respective distributions. Greater emphasis was placed upon Taylor's scale than Winne's because of its higher reliability and its more extensive use in previous experimental studies of anxiety [10]. One hundred and twenty-eight women, 64 from each criterion group, served as subjects in the experiment.

Achromatic copies of Rorschach Cards II, III, VIII, and IX were carefully prepared by a special photographic process to insure faithful reproduction of the inkblots. These particular cards were chosen for the following reasons. First, Cards II and III represent one kind of color application while cards VIII and IX represent another. The former have but a single color, bright red, while the latter have a variety of colors blending together. It is quite possible that one type of color would have a different effect than the other. Second, Cards II and III are only partly colored and partly black and white, whereas Cards VIII and IX are completely colored. The advantages of using these two pairs of cards rather than any others will become apparent when the experimental design has been discussed.

Since it was desirable to study simultaneously a number of independent variables in order to isolate interactions as well as main effects, a rather elaborate experimental design was employed. Thirty-two basic plans were constructed in which stimulus configuration or inkblot, color, and order of presentation were varied systematically in a completely balanced fashion. One half of the subjects received a colored card on the first trial, the other half getting an achromatic card first. Similarly, each stimulus configuration appeared equally often on each trial. A standard procedure with full inquiry, according to Klopfer and Kelley [11], was used; each individual was administered all four inkblots, two in the original Rorschach version and two in the achromatic form.

The number of possible permutations or

---

[1] Formerly at The University of Texas.

orders of presentation of the four inkblots, ignoring the presence or absence of color, is 24. This was reduced to eight by pairing Card II with Card III and Card VIII with Card IX. Thus, both members of a given pair always appeared together in either the first or last two trials. When the color variable (chromatic vs. achromatic) was superimposed on stimulus configuration, a total of 32 plans was required to balance the design completely.

With 128 subjects, four replications of the basic design were possible. Two additional independent variables, Examiner X vs. Examiner Y, and high-anxiety group vs. low-anxiety group, used up the replications. Since much of the confusion in the Rorschach literature may be a function of examiner variance, two examiners were used, one with considerable experience in administering and interpreting the Rorschach and the other with formal training on the Rorschach but little clinical experience. Subjects within each anxiety group were assigned at random to examiner and plan by a third person, the only individual who knew whether a particular subject belonged to the high- or low-anxiety group. This was necessary to avoid the possibility that the examiner's administration of the cards might be influenced by knowledge of the subject's classification.

Every Rorschach response was scored for location, form level, movement, shading, and content. Because of the subjectivity necessarily involved in scoring form level, each response was scored twice according to Beck's tables for $F + \%$ [1] by two psychologists working independently. The interscorer reliability for $F + \%$ was .83. The actual scoring reliability was probably higher since every scoring discrepancy was resolved by a conference of all three authors. The number of responses, the reaction time, and any overt remarks by the subject were also recorded for each card. In addition, every protocol was scored according to Elizur's signs of anxiety [8], a list of signs based on content analysis.

Results

Since all possible advance clues to the subject's classification were withheld from the examiners, the more experienced of the two examiners (X) judged each of his 64 subjects as "anxious" or "nonanxious" solely on the basis of behavior during the individual testing session. No judgments were made by the other examiner since he did not feel sufficiently competent to attempt this task. A number of signs were employed by X in making a judgment: the degree of composure shown by the subject upon entering the room, questions regarding the purpose of the test, the subject's attitude toward X, his readiness to cooperate, and his attentiveness to directions. After administering the test, if X still felt that he needed more information before making a final judgment, he conducted a short interview in which attitudes toward the university, school grades, social activities, and plans after graduation were discussed. The basis of judgment could best be described as intuitive and global, consisting primarily of behavioral signs other than specific responses to the test materials. A comparison of the judgments made by X with the actual criterion grouping is presented in Table 1. That he was able to detect the high- and low-anxiety groups with such a high degree of accuracy attests to the validity of the anxiety criterion.

Table 1

Success of Examiner X in Judging Anxiety Solely on the Basis of Subject's Behavior during the Testing Period

Examiner's judgment	Grouping based on Taylor's scale		
	High anxiety	Low anxiety	Total
High	26	9	35
Low	6	23	29
Total	32	32	64

$\chi^2 = 16.14$ ($p < .001$).

Three general levels of analysis were made: one based upon response to the first card only in the series of four cards presented to each subject; one dealing with only the first two cards in the series; and one concerned with data such as total number of responses and $F + \%$ arising from all four cards.

When only the first card is considered, four of the primary variables—examiner, anxiety, color, and Card II vs. Card III—can be arranged in a 2^4 factorial design with four replications to provide an adequate error term. A similar design exists for Cards VIII and IX.

Ignoring the differences cited earlier between these two pairs of cards would allow combination of both designs with double the number of cases for evaluating the effects of examiner, anxiety, color and their interactions.

The dependent variables or test scores that were analyzed according to this design consisted of the following: (*a*) reaction time, the time from initial exposure of the card until the first response; (*b*) number of responses; (*c*) number of responses using the whole inkblot; (*d*) $F + \%$; (*e*) form level of the first response only; (*f*) number of responses having some kind of movement as a determinant; and (*g*) number of responses with shading as a determinant.

Two general kinds of statistical procedure were used throughout the study. When the test scores were fairly continuous and normal (or could be readily transformed into a normal distribution), analysis of variance methods were employed. At the first level of analysis this proved to be the case, only for reaction time. When the test scores were categorical in nature or relatively discrete and severely skewed, the distributions were dichotomized near the median and handled by chi square or Fisher's exact method [6] for testing the significance of the difference between two percentages.

The second level of analysis dealt with the first two trials together rather than Trial 1 alone. The use of a two-by-two latin square combined with a factorial design made possible an analysis of the effect of order of presentation of the color (colored-before-achromatic vs. achromatic-before-colored) as well as practice effects, anxiety, color, and the interactions of anxiety with color, trial, and order. This was done separately for Cards II and III and for Cards VIII and IX. Differences among the stimulus configurations were completely lost owing to confounding. This design was employed for the analysis of reaction time only since the other variables were too discrete and skewed.

The third level of analysis was concerned with data obtained from the subject's performance on all four cards. In this case only anxiety and examiner could be studied as independent variables, using a two-by-two factorial design. The test data comprising dependent variables consisted of the following: (*a*) mean reaction time to all four cards; (*b*) difference between mean reaction time to the colored cards and that to the achromatic ones; (*c*) total number of responses; (*d*) total $F + \%$; (*e*) number of cards on which the first response had a poor form level; (*f*) difference between $F + \%$ for the colored cards and $F + \%$ for the achromatic ones; (*g*) number of whole responses; (*h*) movement; (*i*) shading; (*j*) *Sum C* scores; (*k*) Elizur's anxiety score; (*l*) animal content $(A + Ad)$; (*m*) human content $(H + Hd)$; (*n*) rejected cards; and (*o*) spontaneous, immediate remarks concerning the color. The first four of these scores were handled by analysis of variance while the remainder were analyzed using chi square and related procedures.

Reaction time. Much has been said in Rorschach literature concerning the psychological meaning of reaction time. A significant increase in reaction time to colored cards is generally considered an indication of "color shock" [11, 14]. Emotionally disturbed individuals will presumably have a long reaction time or show other disturbances due to colors, especially on Cards II and III. From this we might expect that, in general, mean reaction times for the colored cards would be longer than those for the achromatic versions. Moreover, there should be a significant interaction between color and anxiety with respect to reaction time if the above theory is correct.

A more comprehensive analysis was made of reaction time than of any other variable, mainly because of the continuous nature of reaction time scores, which permits the use of analysis of variance. Since the original scores were skewed in the typical fashion of reaction time data, an inverse square root transformation was employed to normalize the score distributions prior to analysis. An analysis of variance of the transformed reaction time scores for Trial 1 only and for Rorschach Cards II and III is presented in Table 2.

In this analysis, all insignificant, higher order interactions were pooled with the within sum of squares to provide an error term with 57

Table 2

Analysis of Variance of Transformed Reaction Time Scores for Rorschach Cards II and III, First Trial Only

Source of variation	Mean square	*df*	*F*	*p*
Examiner	220.9	1	1.63	—
Anxiety	11.4	1	.08	—
Inkblot	.8	1	.01	—
Color	876.2	1	6.50	.02
Examiner by color	581.1	1	4.31	.05
Examiner by anxiety by inkblot	1095.5	1	8.13	.01
Pooled residual	134.8	57		

df. Only one of the main effects was significant. Reaction time to the colored cards was significantly faster than response to the achromatic cards. Although this was more the case for Examiner Y than for Examiner X, as evidenced by the significant examiner-color interaction, these results suggest, contrary to the opinion of many clinicians, that bright red color, as applied on Cards II and III, evokes an earlier response than if these cards were achromatic.

The critical source of variance from the point of view of Rorschach theory, the interaction between anxiety and color, was negligible. Both "anxious" and "nonanxious" women responded in an identical manner. The hypothesis that differences in reaction time to colored and achromatic inkblots are related to anxiety has not been substantiated by analysis of response to the first card only.

The interpretation of the significant triple interaction, examiner-anxiety-card, is unclear. Assuming that this is a nonchance result, it probably reflects examiner differences more than anything else and would prove theoretically important only if one or more of the lower order effects involving these three variables had also been significant.

The same analysis of Cards VIII and IX revealed no significant factors related to reaction time. Unlike Cards II and III, reaction time to these cards is the same in both the multicolored, standard form and in the achromatic experimental version. These results are consistent with the findings of Dubrovner *et al.* [5], who found no differences in reactivity that could be attributed to color. The present study suggests, however, that the bright red color may stimulate the subject to respond more quickly.

On the other hand, Siipola [16] reported just the opposite for a number of colored areas of the Rorschach inkblots. Using matched colored and achromatic versions of smaller inkblots obtained by cutting out the usual detail areas of Cards II, III, VIII, IX, and X, she found consistently longer mean reaction times for college women when color was present than when it was absent. Additional experimentation, with color and inkblot varied systematically in an extensive factorial design, is necessary to resolve these apparent contradictions.

Table 3 contains the results for reaction time on Cards II and III when the first two cards presented to the subject are considered simultaneously. Only one source of variation, the order of presentation of color, proved significant. The colored-before-achromatic order yielded a significantly faster reaction time than the reverse order of presentation. Again it appears that the bright red colors are stimulating to the subject, this time shortening the latency of his response to the card which follows.

At the third level of analysis, employing all four cards, two scores were derived from the reaction time data: (*a*) mean reaction time; and (*b*) difference between mean reaction time to the colored cards and mean reaction time to the achromatic cards. The latter was of particular interest since previous writers [2, 11] on color shock have stated that a delay in reac-

Table 3

Analysis of Variance of Transformed Reaction Time Scores for Rorschach Cards II and III, First Two Trials

Source of variation	Mean square	*df*	*F*	*p*
Independent observations				
Anxiety	0	1	0.00	
Order of color	2080	1	7.07	.01
Order by anxiety	488	1	1.66	
Residual between subjects	294	60		
Correlated observations				
Trial	7	1	0.10	
Color	66	1	0.96	
Trial by anxiety	6	1	0.09	
Color by anxiety	149	1	2.16	
Residual within subjects	69	60		

tion time to a colored card is one sign of "shock" that presumably occurs with greater frequency in neurotic or anxious individuals than in others. An analysis of variance was made for each variable, breaking down the total variance into components attributable to examiner, anxiety, and examiner-anxiety interaction, each having 1 *df*, and an error term with 127 *df*. In neither case did either of the main effects or their interaction approach significance.

The results of these three different sets of findings concerning the extent to which reaction time is a function of examiner, anxiety, inkblot, color, and their interactions can be summarized as follows. Contrary to the opinion of many psychologists, there is little evidence to support the use of a long reaction time to a chromatic inkblot as a sign of color shock. Likewise, there is little support for the contention that anxious individuals take longer to respond to colored than to black and white inkblots, whereas nonanxious subjects do not. If anything, the bright red color on Cards II and III apparently stimulates most individuals to respond more quickly. This stimulation may carry over to the next card, as illustrated by a quicker response to the second card when the first inkblot was partly covered by bright red color. Throughout the analysis of reaction time a consistent difference was found between Cards II and III on the one hand and Cards VIII and IX on the other. This finding raises some interesting questions for further research concerning the comparative effects upon reaction time of variation in hue, configuration, and mode of applying the color to the inkblot.

Number of responses. Although the number of responses given to the first card ranged from one to fifteen, most women gave only one or two responses per card. Because of the discreteness and skewness of the data when only the first card was considered, it was necessary to resort to less efficient methods than analysis of variance to determine the role played by each independent variable. The distribution of scores was dichotomized close to the median, and fourfold tables were constructed; one independent variable was considered at a time and the others ignored. This was done separately for Cards II and III and for Cards VIII and IX as well as for all four inkblots combined. The only variable which showed a significant relationship to number of responses on the first card was color for Cards II and III. Inspection of Table 4 reveals that significantly more women gave more than one response to the colored cards than to the achromatic ones. Apparently the bright red colors on Cards II and III stimulate the individual to greater productivity and also to respond more quickly than otherwise.

Table 4

Number of Responses to the First Card as a Function of Color, Rorschach Cards II and III

Number of responses	Color of inkblot		Total
	Chromatic	Achromatic	
2 or more	22	13	35
1 or less	10	19	29
Total	32	32	64

$\chi^2 = 4.04$ ($p = .05$).

When the total number of responses for all four cards was considered, the frequency distribution was only moderately skewed. An analysis of variance using a two-by-two factorial design revealed no differences among examiners or high- and low-anxiety groups.

These negative results are contrary to those obtained by Eichler [7] and by Westrope [18]. Eichler reported a significant decrease in the number of responses given by individuals under stress conditions. However, in a study relating Rorschach variables to Taylor's anxiety scale, Westrope found just the opposite. Significantly more responses were given by her high-anxiety group than by the low-anxiety group.

The interaction between examiner and anxiety was also negligible. Since the distribution of scores was moderately skewed, an additional test was made by dichotomizing the distribution at the median (nine responses) and constructing fourfold tables for each independent variable. Again, no differences between groups were revealed. The results obtained in this experiment suggest that there is little relationship between the number of responses and manifest anxiety as measured by the MMPI. Although previous experimenters [4, 5, 13, 15] have concluded that color does not increase productivity, the present findings indicate that the bright red hue on Cards II and III may result in a greater number of responses.

Form level. Each response was scored either "plus" or "minus," using Beck's tables [1] as a guide. When only responses to the first card were considered, it was discovered that no women gave minus responses to Card III, only 2 gave one or more minus responses to Card VIII, 4 to Card II, and 16 to Card IX. Since a balanced design had been used by which each card would be presented first to 32 different women, it is quite clear that Card IX is more likely to yield a minus response than any of the other inkblots used in this study. This is compatible with the often reported fact that Card IX has the longest reaction time and is the most frequently rejected card.

Using presence of one or more minus responses to the first card vs. no minus responses as one variable and one of the independent variables (examiner, anxiety, inkblot, color) as the other, fourfold tables were constructed and tested for significance. With the exception of the differences already noted from one inkblot to the next, none of the relationships even approached statistical significance. Similar results were obtained when $F + \%$ for the first card was analyzed.

Consideration of total $F + \%$ based upon all four cards yielded different results. A summary of the analysis of variance for the two-by-two factorial design is presented in Table 5. There was no appreciable difference in the mean $F + \%$ for the high- and low-anxiety groups. This is consistent with the negative results reported by Westrope (18), who also used criterion groups based on the Taylor anxiety scale, for $F + \%$ on the colored cards alone as well as for all ten cards. However, the subjects tested by Examiner Y gave a significantly greater percentage of responses with poor form level than those tested by Examiner X. Such results, though indicating the importance of examiner differences, are of little theoretical consequence since the examiner-anxiety interaction is negligible.

Table 5

Analysis of Variance for Total $F + \%$

Source of variation	Mean square	*df*	*F*	*p*
Examiner	1378	1	5.58	.02
Anxiety	43	1	0.18	—
Examiner by anxiety	8	1	0.03	—
Residual	247	124		

Two other variables dealing with form level were studied, the number of cards on which the first response had poor form level and the difference between $F + \%$ on the colored and achromatic cards. The latter was of particular interest because of the investigation by Williams [19] suggesting that drop in form level on colored cards is a sign of poor control and may also be related to anxiety. Neither variable proved related to either examiner or anxiety in the present study. Fifty-five per cent of the women in both the high- and low-anxiety groups gave all four first responses with good form level. Both groups also showed the same form level on the colored cards as they did on the achromatic versions. Contrary to the findings of Lazarus [12], there is no evidence to support the hypothesis that a drop in form level generally occurs in moving from a black and white inkblot to a colored one.

Other Rorschach variables. None of the remaining variables was sufficiently continuous in nature to permit the use of analysis of variance. Consequently, it was necessary to resort to categorical treatment, using for the most part fourfold tables to investigate the significance of the major factors under consideration.

The number of responses using the whole inkblot, the number employing movement as a determinant, and the number using shading were analyzed for the first card only as well as for the summary data based on all four cards. In addition, *Sum C,* Elizur's anxiety score, animal and human content, rejected cards, and occurrence of spontaneous, immediate remarks to the color were carefully investigated for any differences between examiners or between the high- and low-anxiety groups. Aside from minor differences between examiners in the number of movement responses, no significant factors were discovered for any of the remaining variables.

The negative results for Elizur's anxiety score are of special interest because of its reported validity as a measure of anxiety [8, 18]. The mean score for the high-anxiety group was 1.26, while that for the low-anxiety group was 1.15, a negligible difference.

One explanation for the lack of significant results for Elizur's anxiety score, as well as for some of the other Rorschach variables, is that the use of only four cards, instead of the standard set of ten cards, results in scores of lower reliability. Reducing the number of cards administered to each subject may have lowered the reliability of Rorschach scores in the present study. On the other hand, the number of subjects employed is sufficiently large that a rather small relationship should prove statistically reliable. Although lack of such a relationship in a single experiment does not prove its nonexistence, it does raise serious questions concerning the uncritical adoption of the Rorschach score in question for use in the clinic.

Summary and Conclusions

This experiment was designed to investigate the relationships of anxiety, color, stimulus configuration, order of presentation, and examiner with a number of different Rorschach variables. High- and low-anxiety groups containing 64 subjects each were selected from 696 college women on the basis of MMPI scores for Taylor's Manifest Anxiety Scale and Winne's Scale of Neuroticism. Achromatic copies of Rorschach Cards II, III, VIII, and IX were carefully prepared and used together with corresponding standard cards. Four inkblots, two colored and two achromatic in a predetermined order and balanced design, were administered in a standard fashion to each subject by one of two examiners. The more experienced of the two examiners judged each subject as "anxious" or "nonanxious" at better than the .001 level of significance, solely on the basis of behavior during the individual testing session; this provided a check on the adequacy of the criterion.

The results and conclusions were as follows:

1. The colored versions of Cards II and III yielded significantly shorter reaction times and significantly more responses than did the achromatic versions of these same inkblots. However, no differences ascribable to color were obtained for these same variables on Cards VIII and IX. This finding suggests that the bright red color of Cards II and III is more stimulating than the multicolored pastel shades of Cards VIII and IX. The shorter reaction times to the colored versions of Cards II and III are contrary to expectations from color shock theory.

2. The colored-before-achromatic order of presentation resulted in a significantly shorter mean reaction time for Cards II and III than did the reverse order, achromatic before colored. This is consistent with the hypothesis that the bright red colors of Cards II and III stimulate the individual to respond more quickly.

3. Women tested by Examiner Y gave a significantly greater percentage of responses with poor form level than did women tested by Examiner X, indicating the importance of controlling the examiner as a variable in research of this type.

4. Contrary to the expectations of earlier Rorschach theory, none of the remaining Rorschach variables proved significantly related to anxiety, color, or order of presentation.

It is clear from the present research that many conceptions of how color, examiner, personality, and order-of-presentation variables affect performance on the Rorschach test will have to be revised considerably. Further studies similar to this one, which compare several factors simultaneously by use of modern experimental design, are needed to resolve apparent contradictions in the literature.

Received January 8, 1954.

References

1. Beck, S. J. *Rorschach's test.* Vol. I. *Basic processes.* (2nd Ed.) New York: Grune & Stratton, 1949.
2. Benton, A. L. The experimental validation of the Rorschach test II. The significance of Rorschach color responses. *Amer. J. Orthopsychiat.,* 1952, **22,** 755-763.
3. Brosin, H. W., & Fromm, E. G. Rorschach and color blindness. *Rorschach Res. Exch.,* 1949, **4,** 39-70.
4. Buker, S. L., & Williams, M. Color as a determinant of responsiveness to Rorschach cards in schizophrenia. *J. consult. Psychol.,* 1951, **15,** 196-202.
5. Dubrovner, R., Von Lackum, W., & Jost, H. A study of the effect of color on productivity and reaction time in the Rorschach test. *J. clin. Psychol.,* 1950, **6,** 331-336.

6. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
7. Eichler, R. M. Experimental stress and alleged Rorschach indices of anxiety. *J. abnorm. soc. Psychol.,* 1951, **46,** 344-355.
8. Elizur, A. Content analysis of the Rorschach with regard to anxiety and hostility. *Rorschach Res. Exch.,* 1949, **13,** 247-284.
9. Hertz, Marguerite R. The Rorschach: thirty years after. In D. Brower & L. Abt (Eds.), *Progress in clinical psychology.* Vol. I, Section I. New York: Grune & Stratton, 1953. Pp. 108-148.
10. Holtzman, W. H., Calvin, A. D., & Bitterman, M. E. New evidence for the validity of Taylor's Manifest Anxiety Scale. *J. abnorm. soc. Psychol.,* 1952, **47,** 853-855.
11. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.
12. Lazarus, R. S. The influence of color on the protocol of the Rorschach test. *J. abnorm. soc. Psychol.,* 1949, **44,** 505-516.
13. Perlman, J. A. Color and the validity of the Rorschach 8-9-10 per cent. *J. consult. Psychol.,* 1951, **15,** 122-126.
14. Rapaport, D. *Diagnostic psychological testing.* Vol. II. Chicago: Year Book Publishers, 1945.
15. Sappenfield, B. R., & Buker, S. L. Validity of the Rorschach 8-9-10 per cent as an indicator of responsiveness to color. *J. consult. Psychol.,* 1949, **13,** 268-271.
16. Siipola, Elsa. The influence of color on reactions to ink blots. *J. Pers.,* 1950, **18,** 358-382.
17. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.,* 1953, **48,** 285-290.
18. Westrope, Martha R. Relations among Rorschach indices, manifest anxiety, and performance under stress. *J. abnorm. soc. Psychol.,* 1953, **48,** 515-524.
19. Williams, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.,* 1947, **11,** 21-29.
20. Winne, J. F. A scale of neuroticism. An adaptation of the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.,* 1951, **7,** 117-122.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Motor and Fantasy Correlates of Rorschach Human Movement Responses

Jerome L. Singer
Franklin D. Roosevelt VA Hospital, Montrose, N.Y.[1]
and Jack Herman
New York University

One of the most unusual and theoretically interesting features of Rorschach's inkblot test is the human movement determinant (*M*). In linking the individual's perception of humans in action on the blots to stable motility, on the one hand, and imaginativeness or inner living, on the other, Rorschach [11] was in essence providing a concrete formulation of a theory of vicarious functioning similar to that presented by a number of writers in diverse fields of psychology and psychoanalysis [4, 6, 9, 18]. A series of investigations has provided data supporting Rorschach's linkage of motor inhibition and *M* responses [7, 14, 15, 19]. There is also evidence that *M* responses are associated with measures of fantasy or imaginative tendencies [1, 2, 8, 12, 13, 16]. Two earlier studies in this series have reported positive correlations between the number of *M*[2] and the ability of *S*s to delay their speed of writing. A negative correlation between *M* and the degree of spontaneous motor activity manifested during a waiting period also emerged [7, 15].

The objectives of the present investigation included a partial replication of earlier findings [15] with another sample and an extension of the behavioral situations to include measures of fantasy as well as motor activity. The specific hypotheses are:

1. Schizophrenic *S*s assigned to a High-*M* group on the basis of their Rorschach records differ significantly from Low-*M* *S*s on a series of tasks measuring motor activity and fantasy behavior or introspectiveness. High-*M* *S*s show greater delaying capacity, less impulsivity, less spontaneous motor response, and higher scores on measures of fantasy. A subhypothesis in this connection was derived from an earlier study [15] in which it was found that waiting-room activity scores varied not only with *M* but with the relationship of *M* and the weighted color responses (*Sum C*). This finding offered striking support to Rorschach's juxtaposition of these two types of response determinants (*M*: *Sum C*) as an index of relative weight of introversive or extratensive response tendencies. It was hypothesized, therefore, that, at least in the case of spontaneous motor activity (waiting-room behavior), the degree of activity varies with the increasing weight of color in the experience balance of the patients.

2. If Rorschach descriptions of the interrelations of motor and fantasy activity are correct, one may anticipate a certain consistency in *S*s' performance from one situation to the next. Individuals with relatively numerous *M* also manifest greater motor inhibition, less spontaneous motor activity, and more fantasy. Specifically, the subtests that tap various aspects of the motor inhibition – motion perception–fantasy dimension are intercorrelated and define a common factor or meet some appropriate standard of interrelatedness. These correlations may not be high because of the numer-

---

[1] Reviewed in the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions do not necessarily reflect the policy or opinion of the Veterans Administration. The authors are indebted to Drs. Harold Wilensky and Jacob Cohen for helpful criticism, and to staff and psychological trainees at this station for assistance in various phases of the study.

[2] The focus in this study was on human movement alone. Previous studies have yielded equivocal results concerning the relationship of animal movement (*FM*) and inanimate movement (*m*) to *M*. This empirical distinction between the types of movement responses merits considerable research exploration.

ous situational, uncontrolled factors that undoubtedly limit the reliability of a Rorschach examination. Some gross consistency of performance is certainly to be expected, however, if Rorschach's formulation has any validity. When Rorschach factors are used in relative isolation in this investigation with little concern for "configurational" features, any significant correlations between M and behavioral situations should be all the more striking as evidence of the essential meaning of the human movement determinant.

Subjects

In order to repeat some of the conditions of the previous study [15], a new sample of 60 male schizophrenic veterans was employed. The Ss received clinically administered Rorschachs which were rescored in consultation by experienced examiners. After administration of the five experimental procedures, the total sample was divided into High- and Low-M subgroups. The criteria derived from examination of Rorschach score distributions for the sample were more than one M (high M) and a *Sum C* of 2.5 or more (high *Sum C*). The High- and Low-M groups did not differ significantly in IQ (Wechsler-Bellevue Scale), age, educational level, or other Rorschach variables except M. Although differences in total responses (R) between the two main groups were larger than in the previous study, most of this difference could be attributed to differences in M and FM, the mean M of the High-M Ss being roughly eight times as great as that of the Low-M group. Somewhat in keeping with a finding reported by Wittenborn [20], the pattern of color responses was somewhat different for the two groups, although not significantly so. Thus FC responses tended to correlate somewhat better with M and R than with the other color determinants.

On the whole the characteristics of this sample did not differ markedly from those of the group of 50 schizophrenics described in the earlier publication. Diagnostic labels, degree of morbidity, or duration of illness did not seem to vary systematically. Again as noted in the earlier investigation, the "dilated" experience-type group (high M: high *Sum C*) proved to be somewhat more generally productive and "interesting," albeit no less disturbed than the other subgroups.

Procedure

Each S participated in five experimental situations in more or less the same order. Occasional changes in sequence were necessitated when patients became disturbed or showed wandering attention. In all cases the examiner (always one of the authors) was unfamiliar with patients' Rorschach scores to minimize any unconscious influence on performance.

1. *Waiting-room activity ratings.* This procedure, modified from a method described by Rickers-Ovsiankina [10], was described in a previous paper [15]. The S was conducted into a waiting room and informed that E would call for him shortly. He was seated near a table on which were placed a variety of objects and games. The S was then observed for 15 minutes by means of a one-way vision screen. All activity and movements were recorded at 30-second intervals. These protocols were subsequently rated by judges on a six-point scale for degree of activity and environmental responsiveness. The scale itself was simple to use and had reasonably high interrater reliability [15]. A patient who sat quite still except for an occasional shifting of weight or puff on a cigarette would obtain in minimal score, while an active individual who paced the floor, manipulated many of the objects on the table, and spoke aloud would obtain a maximum score.

2. *Motor inhibition task.* This was a simple test modified from Downey's Will-Temperament Scale and employed in previous investigations of this series [7, 15]. The S was required to write the phrase "New Jersey Chamber of Commerce" as slowly as he possibly could without stopping motion or lifting the pencil from the page. A sample of writing at normal tempo was obtained prior to the experimental trial to evaluate the extent of S's effort to comply with the instructions. This precaution was necessary because of the fairly disturbed patient group employed. The data here were the times in seconds taken to write the phrase.

3. *Thematic Apperception Test.* Cards 1, 2, 4, 6BM, and 7BM of the TAT were administered under two conditions. Initially each S was asked to describe each card. After com-

pletion of the five descriptions, the cards were presented again with the usual instructions to make up a story. The protocols were scored by means of a modification of Weisskopf's transcendence index [17], a measure of the extent to which content is introduced into *S*'s story which goes beyond the actual material inherent in the stimulus. Additional characters, emotions attributed to characters, mention of past or future events were all scored as transcendent items according to a system of thought units. A total score for each series was obtained. In order to derive a measure of voluntary fantasy the difference between the transcendence scores for the *description* and *story* conditions was employed. While other investigators have employed more global ratings of over-all creativity of themes [13, 16], it was felt here that Weisskopf's transcendence index would be most easily quantified and reliably rated. This test situation was felt to be a good initial approximation of an individual's capacity to engage in imaginative behavior, manipulating situations, persons, and events "in his mind's eye," so to speak. The *story* instructions in effect demand of *S* that he adopt an attitude toward the "possible," as Goldstein might put it, by introducing situations and persons not directly called forth by the card itself. Actual verbal fluency is not sufficient here, although it bears some relation to the task. Thus, a person might present a lengthy, obsessive description of the card and of the specific clothes or position of the characters yet never obtain a single point for transcendence. A story like: "This boy's dreaming of being a great violinist. If his parents help him and his teacher encourages him he'll practice hard day by day and finally on an evening 20 years from now he'll win fame before a Carnegie Hall audience," on the other hand, while relatively short, has a relatively high transcendence score.

4. *Cancellation test*. This is a very simple procedure based on an unpublished test devised by Dr. Seymour Klebanoff. Rows of letters in randomized sequence on a printed sheet are presented to *S* who is required to encircle all *o*'s and *e*'s. He is told to go as rapidly as he can but be as accurate as possible. The extreme simplicity of the task makes it possible to obtain a measure of impulsivity by noting errors and time, since the task makes little intellectual demand on *S*s. This test was included chiefly in an exploratory attempt to observe whether *S*s with low *M* and high *Sum C* would show many errors. Perhaps more complex situations might have been more discriminating, but preliminary research with similar patients revealed that such tasks made considerable demands on the limited attention of schizophrenics.

5. *Guilford's Thinking Introversion Scale* (*Factor T of the* Inventory of Factors S T D C R). The items on this scale are described by Guilford as defining a factor which involves "an inclination to meditative or reflective thinking, philosophizing, analysis of one's self and others, versus an extrovertive orientation" [5, p. 000]. Certainly this description resembles some of the propensities of High-*M* individuals as described in the Rorschach literature. While the inventory was developed with college students and has some of the usual limitations of questionnaires, it was included here in an exploratory way to obtain some measure of the patient's self-perception as a thinking introvert. In the interests of maintaining a reasonable time limit and feasible battery of tests for this schizophrenic group, only the T scale of the inventory was employed. To facilitate administration to this group, items were printed on cards and sorted in the manner of the MMPI. It must be recognized that these deviations from the standardized mode of administration may well alter the test and increase uncontrolled variability.

Results

Table 1 presents pertinent background information, Rorschach scores, and results on the experimental situations for the High- and Low-*M* groups. Inspection of this table reveals that the general trend of the data is in the hypothesized direction. High-*M* *S*s differ from the Low-*M* patients by manifesting longer inhibition times in the writing task, less spontaneous activity in the waiting room, greater transcendence of stimulus content in the five TAT cards, fewer errors in the cancellation task,[3] and greater tendency toward thinking introversion on the Guilford scale. Analysis of variance for the motor inhibition times reveals an F ratio of 9.6, significant beyond the .01 level, with the group differences also significant by the t test beyond .01. Waiting-room scores also yielded an F ratio of 5.19, significant beyond the .01 level, with the group means differing significantly by the t test at the .01 level. Scores on the transcendence index yielded an F ratio of 5.96, significant at the .01 level, and again a t test yielded significance beyond .01 for the difference in group means. Neither the cancellation errors nor the Guilford scores yielded significant t or F values although trends were evident. Results of these tests, therefore, tend on the whole to support the initial hypothesis.

---

[3] The groups did not differ significantly in speed of performance although Low-*M* *S*s tended to work a little more rapidly.

Table 1

Means for High- and Low-*M* Groups on Pertinent Background Data, Rorschach Variables, and Experimental Situations

Matching variables	Groups High-*M* group ($N=28$)	 Low-*M* group ($N=32$)
Background variables		
Age	29.6	31.4
Education (years)	11.7	11.2
IQ (Wechsler-Bellevue)	111.5	108.2
Rorschach Variables		
M	2.6	0.3**
Sum C	2.8	2.7
FM	4.7	2.1
Shading	2.6	2.3
Total responses	25.3	17.3
Experimental situations		
Motor inhibition time	172.21	62.97**
Waiting-room activity score	1.92	3.13**
TAT transcendence index	21.96	12.69**
Cancellation test errors	3.00	4.25
Guilford Factor T scale	39.42	36.25

**Differences between groups significant beyond the .01 level.

As to the specific prediction concerning the spontaneous motor activity in the waiting room, Table 2 reveals a regular increase in activity scores as the groups change from high *M*: low *Sum C* through low *M*: high *Sum C*. This result is strikingly similar in trend and in actual scores to the results obtained with the population in the previous study [15]. This relationship between overt response and the *M*: *Sum C* ratio emerged only for the waiting-room scores.

Table 2

Means of Waiting-Room Activity Ratings for High- and Low-*M* Ss Subdivided by *Sum C* Scores

Group	N	Rorschach variables *M*	 *Sum C*	Waiting-room activity ratings
High *M*: Low *C*	14	2.3	1.2	1.58
High *M*: High *C*	14	3.2	4.9	2.29
Low *M*: Low *C*	18	0.4	1.2	2.89
Low *M*: High *C*	14	0.3	5.2	3.43

The next phase of the analysis of results involved testing the hypothesis concerning the consistency of *S*s from situation to situation. Two methods for ascertaining whether a common factor underlies *M* and the five experimental situations were employed. The first of these was Blomqvist's test of relatedness [3], a nonparametric method which provides a t test of the significance of concomitant variation of two or more tests for a common population. It involves dichotomization of all test scores into high and low and makes no assumptions concerning the distributions. For this reason it is useful with Rorschach data and particularly where the range of scores is quite limited, as in the case of the *M* response in this group of schizophrenics. The technique does not, of course, provide the precise information available from a factor analysis which affords an estimate of the degree of relationship and of the factor loadings of the various subtests.

Applying Blomqvist's test to the distribution of *M*, treated as one of a battery of tests, and to the five experimental situations results in a t of 8.48, significant beyond the .001 level. This finding suggests that the six sets of scores are indeed grossly related. To evaluate the extent to which all experimental situations share in this concomitant variation, Blomqvist's test was applied to the six sets of scores in pairs yielding the matrix presented in association with p values in Table 3. It is apparent from inspection of this table that a common factor must indeed subsume *M* and these seemingly diverse test situations. Only the Guilford scale fails to show uniformly significant *p* values for relatedness with the other tests.

Table 3

Matrix of *p* Values for Paired Experimental Situations Based upon Application of Blomqvist's Test of Relatedness

	1	2	3	4	5	6
1. *M*		<.01	<.01	<.05	<.01	>.01
2. Motor inhibition	<.01		<.01	<.05	<.05	>.20
3. Waiting room	<.01	<.01		<.01	<.05	>.20
4. Transcendence	<.05	<.05	<.01		<.01	>.20
5. Cancellation errors	<.01	<.05	<.05	<.01		<.01
6. Guilford's Factor T	<.01	>.20	>.20	>.20	<.01	

The application of Blomqvist's test would thus appear to offer support to the second general hypothesis of this investigation. The next step in evaluation of the data involved the calculation of product-moment correlations between Rorschach variables and the experimental situations. This procedure was employed with some trepidation because of the unknown quality of Rorschach distributions, the extremely restricted range of *M* and other variables in this sample, the limited *N*, and, finally, the inevitable interrelationship of *M* and the total responses (*R*), which cannot legitimately be eliminated by partial correlation since the two variables are not experimentally independent. In line with Wittenborn's reasoning [20], however, the extraction of a centroid factor by Thurstone's multiple factor analysis was carried out in a general attempt to test the hypothesis of commonality. The advantage of this method is that some indication of extent of association could be obtained. Other Rorschach variables bearing on the common factor might also be expected to emerge.

Calculation of Pearson product-moment *r*'s for IQ, variables on the Rorschach, and the five experimental situations revealed relatively few significant correlations with the exception of the matrix of *M* and three of the experimental situations, motor inhibition, waiting room, and the TAT transcendence index. Other variables like *FM*, the Guilford, and cancellation error scores showed trends in the expected direction, but the correlation coefficients did not reach significance. Extraction of centroid factors for those variables which, following a cluster analysis, showed sufficient significant interrelations revealed one clear factor which had its highest loadings from *M* and the motor inhibition times (Table 4). These findings tend to offer partial support to the second hypothesis. It is hoped that further research with larger groups and normal *S*s who may be expected to show a greater range of *M* responses will provide a more suitable test of this hypothesis using the factor analytic approach.

Table 4

Matrix of Intercorrelations for *M* and Three Experimental Variables

Variable	1	2	3	4
1. *M*		.539	–.270	.347
2. Motor inhibition	.539		–.244	.293
3. Waiting-room activity	–.270	–.244		–.077
4. TAT transcendence index	.347	.293	–.077	
Centroid factor loadings	.749	.714	.376	.465

A final use of correlation was attempted to evaluate the role of the color variables. Although *Sum C* and the separate color scores did not yield any significant relationships with other variables, inspection of Table 2 suggests that at least for waiting-room activity ratings some correlation between the *M* : *Sum C* ratio and experimental situations must exist. Scores from 1 to 4 were assigned to the High-*M* : Low-*Sum C*, High-*M* : High-*Sum C*, Low-*M* : Low-*Sum C*, and Low-*M* : High-*Sum C* groups in that order. Calculation of product-moment *r*'s for this *M* : *Sum C* score and the experimental situations revealed no changes or lowering of their correlations with *M* alone. The correlation for the waiting-room scores and this variable was + .47, almost twice that obtained when *M* alone was involved. Thus there seems to be here a hint of validity to Rorschach's linkage of movement and color, and further exploration with other behavioral correlates clearly seems warranted.

Discussion

The data of this investigation on the whole lend support to the view that a person's capacity or tendency to control overt motor activity is related to his capacity for imaginative or fantasy activity. Rorschach's human movement determinant may indeed be a gross but valid measure of these alternative response tendencies. While specific predictions in the individual case are not warranted from the results presented here, it is clear that this motor inhibition—fantasy dimension merits more extensive exploration. Should further experimentation extend the scope of this triadic relationship of motion perception, motor response, and imagination to other behavioral situations or to other populations, a new approach to understanding an important phase of personality development may be opened. Can it be shown that control of motor response, or, more generally, postponement of gratification is accompanied by increased activity on what Lewin [6] has termed the "level of irreality"? At least in moderation such control has great functional value for man. The ability to defer immediate motor response directed toward need gratification with the consequent resort to the realm of fantasy gives man a control of his future through imagery and planning. Indeed such a pattern of postponement of immediate effort and evolution of a timeless behavioral space available for what Freud [4] has called "experimental action" may well be one of the key phases of maturation and, more broadly, civilization. It may also make possible a medium for experiencing anxiety unknown to animals or primitive man. Rorschach's M response is significant insofar as it serves as a measure of presumably long-standing personality characteristic reflecting this pattern of motor control and enhanced fantasy. Perhaps in a more refined and readily quantifiable form[4] this threshold for perception of humans in motion may serve as a valuable clue to important response tendencies that may differentiate not only pathological groups but diverse subcultures within our society. The "deferred gratification pattern" often represented as characteristic of American middle-class values might well be explored by tests of the type used here or more complex extensions of them. Of interest might be an attempt to link early parental training or perceptions of familial figures to the development of the M response and its behavioral correlates.

Summary and Conclusions

Previous studies in this series had demonstrated significant relationships between Rorschach human movement responses and measures of motor inhibition and spontaneous motor activity, thus affording partial support for Rorschach's postulated relationship of overt motor activity, perception of motion, and fantasy. The present investigation was designed as a partial replication of earlier work and an extension to include measures of fantasy or imaginative capacity as well as motor activity. It was hypothesized that *S*s differing in Rorschach M productivity would differ on tasks tapping motor control and fantasy. It was also anticipated that they would manifest rough consistency in their performance on the motor and fantasy situations in support of the general theory of vicarious motor, sensory, and fantasy functioning.

Sixty male schizophrenic patients who had recently been administered Rorschachs underwent five experimental procedures. The *S*s were divided into High- and Low-M groups on the basis of Rorschach protocols, two M or more being the dividing point. High-M *S*s ($N = 28$) did not differ significantly from Low-M *S*s ($N = 32$) in age, diagnosis, education, IQ, or other pertinent Rorschach variables. The experimental situations were: (*a*) observation of *S*s alone in a waiting room with ratings of the degree of activity manifested; (*b*) a motor inhibition task, ability to write a phrase as slowly as possible; (*c*) Thematic Apperception Test stories scored for transcendence, i.e., introduction of content not actually represented in the stimulus; (*d*) Klebanoff cancellation test scored for errors; (*e*) Guilford's scale for the thinking introversion factor (T) from the *Inventory of Factors STDCR*. Analyses of variance and t tests were employed to evaluate significance for the individual tests. Correlational procedures and Blomqvist's nonparametric test of relatedness were employed to evaluate consistency of performance from test to test.

Results on the whole supported both general hypotheses. High-*M* *S*s differed significantly from Low-*M* *S*s by manifesting longer motor delaying capacity, less spontaneous activity during the waiting period, and more frequent introduction of characters or incidents not directly represented in TAT stimuli. They showed nonsignificant trends to make fewer cancellation errors and to present greater self-description as thinking introverts than did Low-*M* patients. For the waiting-room spontaneous activity, the relationship of Rorschach color and movement proved to be particularly important, replicating a finding in an earlier study. With Blomqvist's test, a significant interrelationship of *M* and the five experimental situations emerged, suggesting the presence of a common factor underlying the six sets of scores. Extraction of a centroid factor from a matrix of those intercorrelations which reached significance revealed fairly high factor loadings for *M*, waiting-room activity scores, motor inhibition, and TAT transcendence. The factor analytic procedure, although based on not very tenable statistical assumptions for this sample and these distributions, also appears to lend at least partial support to the hypothesis that a common dynamic factor underlies motor control, motion perception, and fantasy behavior.

---

[4] A promising attempt in this connection was reported recently by Barron [2].

Received January 14, 1954.

References

1. Barrel, R. Subcategories of Rorschach human movement responses: a classification system and some experimental results. *J. consult. Psychol.*, 1953, **17**, 254-260.
2. Barron, F. Threshold for perception of human movement in inkblots and its personality correlates. *Amer. Psychologist*, 1953, **8**, 317-318. (Abstract)
3. Blomqvist, N. Some tests based on dichotomization. *Ann. math. Statist.*, 1951, **22**, 362-371.
4. Freud, S. *Collected papers.* Vol. IV. London: Hogarth, 1946.
5. Guilford, J. P. *An inventory of factors S T D C R.* Manual (Rev. Ed.). Beverly Hills, Calif.: Sheridan Supply Co., 1940.
6. Lewin, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.
7. Meltzoff, J., Singer, J. L., & Korchin, S. J. Motor inhibition and Rorschach movement responses: a test of the sensory-tonic theory. *J. Pers.*, 1953, **21**, 400-410.
8. Pickering, W. D. A comparison of predominant verbal levels on the Thematic Apperception Test with the Rorschach experience-balance. Unpublished doctor's dissertation, Univer. of Pittsburgh, 1950.
9. Rapaport, D. *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951.
10. Rickers-Ovsiankina, Maria. Studies of the personality structure of schizophrenic individuals. I. *J. gen. Psychol.*, 1937, **16**, 153-178.
11. Rorschach, H. *Psychodiagnostics.* New York: Grune & Stratton, 1942.
12. Schumer, Florence. Some behavioral correlates of Rorschach human movement responses. Unpublished doctor's dissertation, Yale Univer., 1949.
13. Shatin, L. Relationships between the Rorschach test and the Thematic Apperception Test. Unpublished doctor's dissertation, Harvard Univer., 1950-51.
14. Singer, J. L., Meltzoff, J., & Goldman, G. D. Rorschach movement responses following moto[illegible] inhibition and hyperactivity. *J. consult. Psychol.*, 1952, **16**, 359-364.
15. Singer, J. L., & Spohn, H. E. Some behavioral correlates of Rorschach's experience-type. *J. consult. Psychol.*, 1954, **18**, 1-9.
16. Vernier, Claire M., & Kendig, Isabelle V. Analysis of the relationships between various measures of creative productivity in two projective tests. Paper read at Amer. Psychol. Ass., Chicago, 1951.
17. ✓ Weisskopf, Edith A. A transcendence index as a proposed measure in the TAT. *J. Psychol.*, 1950, **29**, 379-390.
18. Werner, H. Motion and motion perception: a theory of vicarious functioning. *J. Psychol.*, 1945, **19**, 317-327.
19. Werner, H. Perceptual behavior of brain-injured mentally-deficient children. *Genet. Psychol. Monogr.*, 1945, **31**, 51-110.
20. Wittenborn, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1950, **14**, 261-267.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

The Ammons FRPV Test and the WISC for Remedial Reading Cases[1]

Louis M. Smith and Arline R. Fillmore
Psycho-Educational Clinic, University of Minnesota

The Ammons Full Range Picture Vocabulary Test and the Wechsler Intelligence Scale for Children were given to 91 children diagnosed as reading disability cases at the Psycho-Educational Clinic. A reading disability was defined as a reading grade level, on a standardized achievement test, at least one year below the mental grade level on either the WISC or the Stanford-Binet. Eight of the subjects were girls and 82 were boys. The mean chronological age was 10.9 years, the mean reading grade was 3.2 years, and the mean intelligence level on the WISC was 101. Most of the children received both tests in the initial diagnostic interview, but some were given the Ammons while they were enrolled in the summer reading school.

Form A and Form B of the Ammons were comparable in mean and standard deviation. When combined the mean was 107.89, the standard deviation was 13.70. The mean of the WISC Verbal Scale was 99.66, Performance Scale 101.87, and Full Scale 100.85. The standard deviations were respectively 10.87, 12.51, and 10.71. The correlation between Form A and Form B of the Ammons was .90. The two forms were combined into a single scale and then correlated with the WISC scales. The correlations were .73 with the Verbal Scale, .54 with the Performance Scale, .75 with the Full Scale, and .61 with the scaled score of the Vocabulary subtest. When the correlation between the combined forms of the Ammons and the WISC Full Scale was corrected for the restriction of range on the WISC, the value was raised to .82.

In conclusion, the present writers feel that the Ammons is a useful screening test of intelligence for children with reading handicaps. The differences in mean scores indicate a possible limitation. However, the full significance of this cannot be fully weighed at the present time. Krugman, *et al.* [1] reported differences between the WISC and the Stanford-Binet similar in magnitude and direction to the present differences between the WISC and the Ammons.

Brief Report
Received June 2, 1954.

Reference

1. Krugman, Judith I., Justman, J., Wrightstone, J. W. & Krugman, M. Pupil functioning on the Stanford-Binet and the Wechsler Intelligence Scale for Children. *J. consult. Psychol.*, 1951, **15**, 475-483.

---

[1] An extended report of this study may be obtained without charge from Louis M. Smith and Arline R. Fillmore, Psycho-Educational Clinic, University of Minnesota, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4321 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

A *P* Scale for the Rorschach: A Methodological Study

Philip H. DuBois and Arnold H. Hilden
Washington University

The present study represents the application of a relatively new method of item analysis [3, 6] to Rorschach data. It can be readily shown that the total variance of a test can be divided into two parts: the sum of the item variances and the sum of all the covariances among the items. The total variance of a test with zero intercorrelations among the items, and hence zero interitem covariances, would be the sum of the item variances. Each item would then have nothing in common with the other items. The test would be completely heterogeneous, measuring different qualities rather than a single quality. It would be impossible to give a name to such a test, unless it predicted a criterion similarly heterogeneous. In that case its unity would come from without rather than from within.

In general, a psychometric test should be designed to measure a unitary characteristic, revealed by positive intercorrelations among the items. If these correlations are too high and approach unity, the test makes too few discriminations to be of value. With the type of material used in the present study, however, this difficulty was not encountered. Our procedure for selecting items was to maximize the saturation, defined as the ratio of the sum of the item covariances to the total variance of the test. This formula may be written as follows:

$$S = \frac{2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} C_{ij}}{\sigma^2_x}$$

$$= \frac{\sigma^2_x - \Sigma\sigma^2_i}{\sigma^2_x} = 1 - \frac{\Sigma\sigma^2_i}{\sigma^2_x},$$

in which C_{ij} represents the covariance between any two items, and σ^2_x is the total variance.

It is readily apparent that maximizing the saturation is essentially the same operation as maximizing the internal consistency as estimated by Kuder-Richardson Formula 20. We can convert a saturation coefficient into r_{tt} by multiplying by $\frac{n}{n-1}$ in which n is the number of items.

It occurred to us that this procedure offered promise as a method of scaling Rorschach variables. Rorschach literature, while voluminous, has not drawn heavily on conventional psychometric methods.

The popular response, *P*, was defined by Rorschach [8, p. 359] as follows: "The vulgar answers represent the share in the collective or common method of sensing anything." Subsequently, a number of authors have expressed various definitions or rationales underlying the popular concept. Hertz [4], Beck [1, 2], Klopfer [5], and Rapaport [7], among others, have proposed different *P* lists. All, including Rorschach, define *P* on a definitely statistical basis. It is one of the few Rorschach categories so defined. There is, however, variation as to the exact statistical criterion. Klopfer says [5, p. 177] that "the term popular refers to . . . responses given most frequently . . . to a particular blot area" and "it seems most reasonable to use the total number of subjects . . . as a basis for determining the frequency with which a concept has been selected." He states that since a frequency tabulation of universal validity is not available, he has substituted a selected sample of popular responses which meet four qualifications: (*a*) they must occur frequently; (*b*) frequently used determinants must be present; (*c*) any "normal" person should be able to accept them; and (*d*) a *D* or *W* must be involved.

Table 1

Saturations and Reliabilities (by Kuder-Richardson Formula 20) of the Several *P* Scales

		Group A		Group B		Combined	
Author	n	S	r_{tt}	S	r_{tt}	S	r_{tt}
Beck	21	.48	.51	.41	.43	.45	.47
Hertz	20	.36	.38	.39	.41	.38	.40
Klopfer	10	.15	.17	.24	.27	.20	.22
Rapaport	13	.24	.26	.35	.38	.30	.32
Saturation	18	.63	.67	.51	.54		

Beck insists that *P* is based on form alone. Rapaport gives no exact criterion, but states that "they must occur in at least every 4 or 5 records" [7, p. 313]. Hertz [4] uses a definite statistical basis. Beck [2] has used the most definite and elaborate basis, and has made the most exhaustive attempt to arrive at criteria for establishment of *P* responses.

In the present study the lists proposed by these four authors were taken as representative of populars used in clinical practice in this country. The four lists were combined into a composite list of 47 items, many of which were alternates or variations of others on the list. While the four authors are not in complete agreement on any item, the list was drawn up so that each of the four authors was fully represented.

Subjects

The subjects consisted of 200 normal adults, normal in the sense that they were not clinic patients. Most of them were tested by graduate students enrolled in a projective technique course, while the others were tested by one of the authors.

The population, consisting of 100 males and 100 females, was randomly divided into two groups, Group A, an experimental group, and Group B, a cross-validation group. Mean ages were 31.6 and 30.4 years, respectively.

Procedure

For Group A, the entire matrix of item variances and covariances for 47 items was computed. Saturation and internal consistency reliability coefficients are reported in Table 1 for the scales proposed by the four authors, in the two groups of 100 each and for the combined group of 200. For these scales, variation in saturation and reliability from Group A to Group B comes merely from sampling, since no maximizing is involved. Of the four scales proposed, Beck's has the highest internal consistency, .47, Klopfer's the lowest, .22. Even with Beck's scale, however, the internal consistency is lower than we would wish to have in a good psychometric device measuring a quantifiable aspect of human behavior along a continuum.

In order to develop a homogeneous scale of Rorschach populars, the entire collection of 47 items was considered as a source of possibilities. However, a number of the items, as indicated in Table 2, were considered as overlapping, representing definitions by different authors of the same general idea. Accordingly, only one of the items in such a group of "overlapping" was considered for inclusion at any one time.

A nucleus of four items having the highest available interitem covariances was selected as a starting point. This nucleus consisted of items 1, 13, 19, and 23, as listed in Table 2. At each stage in the development of the saturation scale, all items were tested for inclusion in or exclusion from the nucleus.

It has been shown [6] that any item, *j*, not included in the nucleus, will add to the saturation if

$$\frac{2C_{xj} + \sigma^2_j}{\sigma^2_j} > \frac{\sigma^2_x}{\Sigma\sigma^2_i}.$$

On the other hand any item, *k*, included in the nucleus, should be removed if

$$\frac{\sigma^2_x}{\Sigma\sigma^2_i} > \frac{2C_{xk} - \sigma^2_k}{\sigma^2_k}.$$

A peculiar problem was encountered in the construction of the present scale in the matter of alternative possibilities, that is, different definitions of the same general idea. Accordingly, a formula was developed to determine whether any specific interchange would add to the saturation.

If *j* is any item in the nucleus and *k* is any item outside the nucleus, the saturation will be increased if the following inequality holds:

$$\sigma^2_x\,(\sigma^2_j - \sigma^2_k) > \Sigma\sigma^2_i\,(2C_{jk} + 2C_{xj} - 2C_{xk} - \sigma^2_j - \sigma^2_k).$$

Table 2

Composite *P* List

Item no.	Percentage passed** A	Percentage passed** B	Card and location†	Source‡	Content of the popular responses
1§	58	55	I, *W*	BHKR	Bat, butterfly, moth (R omits moth)
2	4	1	I, *W*	H	Bird (also item 1)
3	2	5	I, *W*	K	Winged creature (also item 1)
4	61	61	I, *W*	K	Items 1, 2, and 3 combined (generic concept)
5	1	0	I, *W*	H	Face of a person (not animal or cat)
6*	10	11	I, *D*3	B	Lower part of body (*D*4 if *D*3 attended to)
7	18	21	II, *W*, ½	BH R	Human being
8	0	1	II, *W*, ½	R	Bear, dog
9*	31	33	II, *D*1	B R	Bear, dog
10	32	40	II, *D*1, 7	H	Bear, dog, cow, animal
11	33	35	II, *D*1	K	Bear, dog, rabbit, bull, rhinoceros variety
12*	17	17	II, *D*3	B	Butterfly, moth
13*	68	83	III, *D*1,9	B R	Human beings (with or without *M*)
14	66	79	III, *D*1,9	HK	Human beings (only if *M* is present)
15	21	20	III, *D*3	B	Bow, bow-tie, hair ribbon
16¶	32	36	III, *D*3	HKR	Butterfly, or item 15
17	46	35	IV, *W*	BH R	Animal skin
18	7	17	IV, *D*2,6	H	Boot, foot, shoe
19*	77	86	V, *W*	BHKR	Bat, butterfly, moth (R omits moth)
20	4	6	V, *W*	H	Bird (also item 20)
21	90	92	V, *W*	K	Winged creature, or items 19 or 20
22*	12	17	V, *D*1	B	Leg, animal or human
23	51	65	VI, *W*, *D*1	BH R	Animal skin (*F* only or with shading)
24*	37	46	VI, *W*, *D*1	K	Animal skin, only when shading is present
25*	51	51	VII, *D*1	BH R	Face or head of human (also if *D*2, ½, *W* used)
26	69	67	VIII, *D*1	B	Animal (Beck's list)
27	89	83	VIII, *D*1	H	Animal (Hertz's list)
28	51	50	VIII, *D*1	K	Four-legged animal, with *FM*
29	93	90	VIII, *D*1	R	Animal
30*	21	24	VIII, *D*3	BH	Vertebrae, ribs, bone structure
31	6	9	VIII, *D*4	BH	Tree, bush
32*	19	20	IX, *D*3	B R	Human being
33	14	17	IX, *D*4	B	Face, head; of a man
34*	16	17	IX, *D*4	H	Face, head; person, not female
35*	16	20	IX, *D*2	H	Face, head; of animal (Hertz's list)
36	1	3	IX, *D*9	H	Tree
37*	52	41	X, *D*1	B	Spider, crab, lobster, crustacean
38	54	39	X, *D*1	H	Spider, crab, octopus, crustacean
39	33	25	X, *D*1	R	Crab, octopus
40	58	46	X, *D*1	K	Many-legged animal, or items 37, 38, 39
41*	17	14	X, *D*2	B	Dog
42‖	21	24	X, *D*4		Worm, caterpillar, serpent (*F* or *FC*)
43*	26	34	X, *D*4	H	Seahorse, or item 42
44	18	20	X, *D*4	K	Elongated animal, with use of color
45*	21	29	X, *D*5	BH R	Rabbit head
46	24	33	X, *D*5	K	Animal head with ears or horns, or item 45
47*	9	11	X, *D*12	B	Sheep

---

* Items remaining in the maximum saturation scale.

** Each group consists of 100 cases.

† Under "Location," numbers refer to Beck's **D** notation.

‡ Capital letters refer to Beck, Hertz, Klopfer, and Rapaport, respectively.

§ Parentheses indicate overlapping items.

¶ Item 16 not correct as to Hertz; noted too late to change.

‖ Item 42 belongs to no author; inserted to test with or without color.

Of the 47 items considered, 18 yielded a saturation of .63 in the experimental group. This was the maximum saturation possible, using no more than one of any group of "overlapping" items. These 18 items are designated with asterisks in Table 2.

Since a maximizing procedure was employed, the true characteristics of the *P* scale so developed should be estimated from Group B. Here the saturation is .51 and the reliability as estimated by Kuder-Richardson Formula 20 is .54. This was the highest internal consistency measure found for any of these scales, but obviously leaves much to be desired.

Table 3

Intercorrelations of *P* Scales ($N = 200$)

Author	*n*	Hertz	Klopfer	Rapaport	Saturation	*M*	*σ*
Beck	21	.88	.71	.86	.88	7.06	2.62
Hertz	20		.76	.82	.83	7.12	2.22
Klopfer	10			.68	.73	4.85	1.61
Rapaport	13				.71	5.80	1.81
Saturation	18					5.61	2.73

The intercorrelations of the five *P* scales for the entire group of 200 cases are given in Table 3. All have a great deal in common, as shown by the intercorrelations, none of which is lower than .68. High intercorrelations are to be expected because of overlap in rationale of the items and in actual item content. The saturation scale correlated .88 with Beck's, with which it has much content in common. Identical content makes all these intercorrelations high. It is of interest that in all but one of the items included in the saturation scale the determinant was form only. This may have some bearing on the point of view put forth by some authorities that determinants other than form should sometimes be considered in the selections of populars. In this connection, the question sometimes occurs as to whether the responses included as a popular should be defined quite specifically or in rather generic terms. For example, in Card 1 for the whole card, should one accept merely the bat, the butterfly, and the moth, but not a bird or a winged creature as a generic term? There are a number of points for which such a question could well arise. This point was tested during the investigation by replacing more specific terms by the more generic ones. In no instance was the saturation affected by more than a point or two.

Conclusions

Of the lists of populars proposed by four authors, Beck's scale exhibits the highest internal consistency. Its reliability, however, is definitely lower than that of acceptable psychometric devices. A revised list of 18 items, developed during the present investigation, is a slight improvement over Beck.

The present study indicates that orthodox psychometric methods can be used with projective material and will aid in the improvement of posited variables. The authors of this paper, a practicing clinician and a worker with conventional tests, were both surprised by the outcome. The internal consistency of the *P* scale, as developed by saturation analysis, was neither as high as the clinician expected nor as low as the psychometrician believed.

Variables which have a high degree of internal consistency may or may not be useful. Variables which have neither internal consistency nor external validity are certainly useless. It would appear that further examination of Rorschach variables is indicated.

Received February 22, 1954.

References

1. Beck, S. J. *Rorschach's test.* Vol. I. *Basic processes.* (2nd Ed., Rev.) New York: Grune & Stratton, 1949.
2. Beck, S. J., *et al.* The normal personality as projected in the Rorschach test. *J. Psychol.*, 1950, **30**, 241-298.
3. DuBois, P. H., Loevinger, Jane, & Gleser, Goldine C. The construction of homogeneous keys for a biographical inventory. *Hum. Resour. Res. Cent., Res. Bull.*, 1952, No. 52-18.
4. Hertz, Marguerite R. *Frequency tables for scoring responses to the Rorschach inkblot test.* (3rd Ed.) Cleveland: Western Reserve Univer. Press, 1951.
5. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.
6. Loevinger, Jane, Gleser, Goldine C., & DuBois, P. H. Maximizing the discriminating power of a multiple score test. *Psychometrika*, 1953, **18**, 309-317.
7. Rapaport, D. *Diagnostic psychological testing.* Vol. II. Chicago: Year Book Publishers, 1946.
8. Rorschach, H., & Oberholzer, E. The application of the interpretation of form to psychoanalysis. *J. nerv. ment. Dis.*, 1924, **60**, 225-248, 359-379.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Sex, Age, and Iconicity as Factors in Projective Film Tests[1]

Charles J. McIntyre
The Instructional Film Research Program[2]
The Pennsylvania State University

This study was undertaken to investigate the probable effectiveness of sound motion pictures as a medium for projective personality testing, and to attempt to define some of the characteristics which should be built into such a test.

Previous studies [2, 4, 5] have indicated that specially produced films may provide a powerful new tool for personality assessment. No attempt was made in these studies, however, to demonstrate that films might be superior to any conventional or widely used projective test, the Thematic Apperception Test (TAT), for example. Such a demonstration appears to be necessary if one is to justify the cost of producing a new test. Nor has any systematic attempt been made previously to indicate what characteristics such films should have in order to be maximally effective.

A related consideration involves the desirability of developing a criterion measure which not only would permit evaluating the limited variables tested in this study, but which could serve as a relatively quick and simple method for evaluating each of many stimulus characteristics that should be evaluated before constructing a motion picture projective test.

Since the success of picture projective tests appears to be dependent upon the degree to which subjects' (*S*s') interpretations of the stimuli are projections of their own personalities, the variables were evaluated in terms of the relative degree to which they permitted or encouraged projection onto the central figure (hero, protagonist).

The emphasis, which appears again and again in TAT literature, upon making the central figure like *S* with respect to age, sex, color, body-build, etc., in order to increase projection is directed to the end of making it possible for *S* to perceive the protagonist as being like himself. In fact, this appears to be the central meaning of projection in this context; namely, that an *S* has projected to the extent that his reported perception of the protagonist agrees with his perception of himself. This is the operational definition of projection used throughout this study.

The following hypotheses were tested.

1. There is a greater than chance agreement between an *S*'s perception of himself and his perception of an ambiguous film protagonist.

This hypothesis means, in terms of the operational definition, that persons viewing an ambiguous film will project their own self-perceptions onto the protagonist.

2. Provided that the total picture remains ambiguous, the amount of projection elicited by a protagonist will increase as the iconicity of the protagonist increases.

Iconicity is a term borrowed from semantics. It is used here to mean something akin to lifelikeness or realism. Apparently, the peculiar advantage of pictures, and especially motion pictures, over other forms of representational art is the great realism with which persons or things can be portrayed.

3. The *S*s will project more onto protagonists of the same sex than onto protagonists of the opposite sex.

4. The *S*s will project more onto protago-

---

[1] This paper represents the substance of a dissertation submitted in partial fulfillment of the requirements for the Ph. D. degree at the Pennsylvania State University.

[2] This research was carried out under contract No. N6onr-269, T. O. VII between the Instructional Film Research Program, The Pennsylvania State University, and the Special Devices Center, Office of Naval Research.

nists of the same apparent age as the *S*s' than onto protagonists whose apparent age is different from the *S*s'.

Hypotheses 2, 3, 4 are all based upon the assumption that increased subject-protagonist similarity will result in increased projection. In hypothesis 2 the element of similarity is, of course, realism or lifelikeness. It says, in effect, that the probability for projection occurring is increased as the protagonists become more like real people and less like paintings or still pictures of people.

The assumption that increased similarity results in increased projection permeates the TAT literature. It was made specifically by Murray [7], Thompson [12], and Symonds [11]. It is implicit in the studies of Crandall [1], and Heppell and Raimy [6] when, in selecting pictures to elicit information in specific areas, they were careful to stipulate that at least one character in their pictures be of the same sex and age as the persons to be tested. In addition, among the rules set forth by Rotter for determining onto which figure an *S* is projecting, one finds, "The subject is likely to identify with a character of the same sex, either one of his own age or a previous age. For example, a young male adult might identify with a young male adult or a boy in the story but would be unlikely to identify himself with an older person, or a female of any age" [9, p. 82]. "Identify" in this context apparently means to project one's own characteristics onto a particular figure in the card.

It should be noted, however, that this assumption is not supported by the findings of Riess, Schwartz, and Cottingham [8, 10] and Weisskopf and Dunlevy [13, 14] which, while they do not indicate what conditions are necessary for eliciting maximum projection, do indicate that, whatever the necessary conditions may be, they are more subtle than simple physical similarity.

Procedures

The films. In all treatments, the stimulus materials were 16-mm. motion pictures which from the beginning of the fade-in to the end of the fade-out were each 35 seconds long. Five film versions were produced for each of five scenes, plus an additional film which was shown at the beginning of each experimental session as a practice exercise. Thus, there was a total of twenty-five different experimental films plus one practice film.

The five experimental scenes upon which the versions were built were selected from five TAT cards so as to embody the sex and age variables of the study. Since it was essential to duplicate as nearly as possible these scenes in the motion pictures, the selection of cards was limited, in that it was necessary to find persons who looked like the TAT characters they were portraying, and to find locations for shooting which were like those of the cards. An additional limiting factor was the necessity of selecting scenes in which the pose was relatively static rather than one implying action. The action versions of the scene had to conform to the general principle of ambiguity, and it was found in planning the scripts that any action coming before or after a dynamic still pose tended to define the meaning of the scene. When a pose was essentially static it was relatively easy to get an actor into or out of that pose without giving definition to the scene. With these limitations in mind, the following scenes were selected as the variables to test the age and sex hypotheses:

Scene 1. Variable: Young boy (TAT Card 1) A young boy is contemplating a violin which rests on a table in front of him.

Scene 2. Variable: Young man (TAT Card 20) The dimly illumined figure of a man (or woman) leaning against a lamppost in the dead of night.

Eron [3] in his normative study of the TAT found that 20 per cent of nonhospitalized *S*s misidentified the figure in card 20. To avoid contaminating the data because of misidentification, *S*s who were tested with this scene in the present study were required to indicate the sex and approximate age of the figure. Those who did not identify the figure as that of a young man were eliminated in the statistical analysis.

Scene 3. Variable: Young woman (TAT Card 8-GF) A young woman sits with her chin in her hand looking off into space.

Scene 4. Variable: Old man (TAT Card 7BM) A gray-haired man is looking at a younger man who is sullenly staring into space.

Although there are two persons in this scene, *S*s were told to make their judgments about the old man only.

Scene 5. Variable: Old woman (TAT Card 6BM) A short elderly woman stands with her back turned to a tall young man. The latter is looking downward with a perplexed expression.

As in Scene 4, there are two persons in this scene. The *S*s were told to make their judgments about the old woman only.

Scene 6. The practice scene shown to all groups (TAT Card 14). The silhouette of a young man (or woman) against a bright window. The rest of the picture is totally black.

The variable of iconicity was built into the five motion picture versions of each of the five scenes. Iconicity has been defined as lifelikeness or realism. In the development of the versions, an attempt was made to keep each successive version of a particular scene as nearly identical as possible, while adding elements that would make the characters more real or more lifelike than in the preceding version. Thus, Version 1, a photograph of a TAT card, was the least iconic since the figures in the card were merely drawn representations of people. The figures in Version 2 were more lifelike in that they were photographs of real people rather than drawings. In Version 3, the characters moved into the scene and thus were endowed with another attribute of real people, i.e., motion. Version 4 extended the principle of Version 3, that is, the characters manipulated and interacted with their environment. Version 5 attempted to add another dimension of behavior associated with people, namely, speech. Since the speech in Version 5 was made unintelligible in order to retain ambiguity, it may be that this version was actually less lifelike than the other versions rather than more, as intended. However, it can be seen that the intent in each case was to add increments of lifelikeness to each succeeding version in order to make the characters more like real, living people than in the preceding versions.

The measure of projection. For the purpose of this study projection has been operationally defined as the degree to which an individual's perception of a film protagonist agrees with his perception of himself. The Minnesota Multiphasic Personality Inventory (MMPI) was used as a pool from which items were drawn to sample *S*s' perception of the protagonist and their perception of themselves. That is, after the film sequence was shown, *S*s were asked to answer true or false to 26 questions which read, for example, "This is the sort of person who feels more intensely than most people do." After responding to these perception-of-protagonist questions, *S*s then took an MMPI for themselves during the course of which they would answer an item which read, "I think that I feel more intensely than most people do." When an *S*'s answer to the question with respect to the protagonist was the same as his answer with respect to himself, this was scored as a "unit of projection" and the sum of these was his "projection" score.

The full MMPI, rather than just the pertinent 26 items, was administered in order to obscure from *S*s the relationship between the items in the film test and the items in the inventory.

The items of the perception-of-protagonist test (the film test) were selected from the MMPI by two persons working independently. Only those items were used which both agreed met the following criteria:

1. The item should not be objectively answerable from the film. That is, there should be nothing about the appearance or action of the protagonist which would suggest a "correct" answer.
2. An item should not have a distinctly "abnormal" answer. Thus an item such as "Someone has control of my mind" although perhaps suitable according to criterion 1, would be unsuitable, since most people would answer this false for themselves and false for the protagonist merely by inference and thus, by the measure, be "projecting" with respect to this item.
3. In the interest of subject cooperation and motivation an item should have the appearance of being answerable from the film. An item such as, "My hands and feet are usually warm enough" would be rejected since an *S* would be likely to feel that he could not reasonably be expected to answer this item with respect to a film protagonist.
4. Items should be drawn from various scales of the personality inventory in approximately equal number.

The numbers of the 26 MMPI items used with each of the sequences were: 8, 13, 26, 32, 39, 54, 64, 79, 82, 91, 93, 94, 102, 165, 171, 242, 257, 267, 292, 299, 336, 362, 371, 399, 407, 416.

The subjects. The *S*s were 425 college students in elementary courses in education, psychology, and sociology. They were assigned randomly to the treatments. Each *S* saw only one of the twenty-five different sequences. Hence, the total subject pool was divided randomly into 25 groups of 7 men and 10 women each.

Test administration. The screen image was held constant at 58½ inches wide and 44 inches high. The arbitrary image size was dictated by the size of the smallest room (and therefore minimum projector-to-screen distance) available for the experiment.

In summary, the *S*s were told that it is well accepted that one of the chief characteristics of motion pictures is their ability to portray complex human emotions, feelings, and interactions, but that the ability of the actors is only one aspect of this. This experiment, they were told, was being conducted to find out about another aspect, namely, the ability of the audience to understand and interpret what they had seen. Hence, the *S*s were told they would see a very brief scene and would then be asked to make certain judgment about it. They were urged to guess when they could not be sure of their answers.

Following the instructions, the practice scene was shown. The *S*s then had 3 minutes to answer 20 questions written by *E* so as to be relatively easy to answer in order to encourage *S*s to think that the task was not really difficult. The same scene was then shown again and the *S*s given a short period to complete answering the questions. The *S*s did not know that this was a practice exercise. Following this the experimental film was shown, *S*s were given 3 minutes to answer the 26 selected items, the same film was shown again, and *S*s were given another brief period to complete answering the questions for that film. After the test papers were collected, it was explained that *E* was interested in relating their judgments of the film characters to certain aspects of personality and that, therefore, they were being asked to complete a personality inventory. The *S*s were permitted to work at their own speed and leave when they had finished.

Results

The experiment has the form of a 5 × 5 × 2 factorial design. There were five scenes, each varying five ways (versions) with the *S*s grouped by sex. Analysis of variance was the principal statistical technique. The data were tested, and it was shown that the conditions of normally distributed scores and homogeneity of variance, necessary in the analysis of variance, were met.

Table 1

Analysis of Variance of Projection Scores

Source of variation	*df*	Mean square	*F*
Between versions	4	57.3705	1.34
Between scenes	4	341.3176	7.98**
Between sexes	1	2.7880	—
Version × scene	16	54.1073	1.65
Version × sex	4	25.7330	—
Scene × sex	4	54.7306	1.67
Version × scene × sex	16	32.7392	1.77*
Within groups	375	18.5386	
Total	424		

* Indicates significance at 5% level.
** Indicates significance at 1% level.

The analysis of variance is summarized in Table 1.

It will be seen in Table 1 that the differences between the means of the scenes were highly significant. It was therefore appropriate to determine the nature and extent of these differences by means of *t* tests. These results are summarized in Table 2.

Table 2

Differences between Means of Scenes

Mean	of Scene	Mean	of Scene	Difference
12.86	1	11.78	2	1.08
12.86	1	10.93	3	1.93**
12.86	1	16.19	4	—3.33**
12.86	1	13.31	5	— .45
11.78	2	10.93	3	.85
11.78	2	16.19	4	—4.41**
11.78	2	13.31	5	—1.53*
10.93	3	16.19	4	—5.26**
10.93	3	13.31	5	—2.38**
16.19	4	13.31	5	2.88**

* Indicates significance at 5% level.
** Indicates significance at the 1% level.

It will be seen in Table 2 that the mean projection score for Scene 4 (old man) was significantly greater than any other scene, and the mean score for Scene 5 (old woman) was significantly greater than the mean of Scenes 2 and 3.

Because of the nature of the measure of projection the need for additional analysis was indicated. Since an individual's projection score was the sum of the times he answered an item the same way on each of two tests, merely by chance an *S* might be said to be "projecting" with respect to a particular item 50 per cent of the time and thus might be expected to get a "projection" score of 13 (26/2) by chance alone. The mean projection score over all *S*s was 13.01 Because the obtained and chance mean scores were virtually identical, chi-square values were computed for the agreement between answers to each item in both tests. The frequency of obtained chi squares for each of the 26 items was compared to the chance distribution of chi square. The resulting chi-square value yielded a p of approximately .15. Hence, the observed distribution of chi square for the 26 items did not differ significantly from the distribution that would be expected to arise if only chance were operating.

The implications of these findings in terms of the specific hypotheses are discussed below.

1. Hypothesis 1 predicted that *S*s would project their own characteristics onto an ambiguous film protagonist. The last finding reported, i.e., that the frequency of agreement between the same items on the two tests (which was the measure of projection) did not differ significantly, seems to indicate, if considered alone, that projection as defined and measured in this experiment did not occur. On the other hand, the *F* ratio for differences between scenes was significant well beyond the 1 per cent level.

It could be true, of course, that the present experiment was the "less-than-one-time-out-of-a-hundred" in which the differences occurred by chance. By definition, this event is unlikely. A more likely interpretation of this apparent contradiction is that no measurable degree of projection occurred with respect to Scenes 1, 2, and 3 but that *S*s did project their own characteristics onto the protagonists of Scene 4 and, possibly, Scene 5. However, while there was sufficient nonchance agreement between the film test and MMPI items to raise the mean projection score on Scene 4 significantly above all other scenes, and on Scene 5 above all but one other scene, these nonchance agreements, when grouped with the apparently chance scores of the other scenes, were insufficient to produce chi-square values which differed significantly from chance.

It, therefore, can be said that *S*s were found to project onto some ambiguous film protagonists, but the result did not indicate that the phenomenon was general.

2. Hypothesis 2 predicted that projection would increase with the iconicity of the protagonist. The iconicity variable was embodied in the versions. The between-versions *F* of 1.34 in Table 1 indicated that no significant differences between the means of versions were found. In view of the fact that the scores of at least three of the five scenes did not differ significantly from chance, it was to be expected that the versions would not differ. When Scenes 4 and 5, in which there appears to have been projection not attributable to chance, were analyzed apart from the other scenes, no significant differences between the means of the versions of those scenes were found. Hypothesis 2 was not confirmed.

3. The third hypothesis predicted that *S*s would project more onto protagonists of the same sex than onto protagonists of the opposite sex. In Table 1, it will be seen that the scene × sex interaction term was not significant. In fact, both men and women projected most onto the protagonist of Scene 4, an old man. Hence, this hypothesis was not confirmed.

4. Hypothesis 4 predicted that *S*s would project most onto protagonists of the same apparent age as themselves. Hence the greatest projection should have occurred onto Scene 2 (young man) and Scene 3 (young woman). It will be seen in Table 2 that these two scenes were the least effective in eliciting projection. The most effective scenes were 4 and 5 in which the protagonists were an old man and old woman respectively. Hypothesis 4 was not confirmed.

Discussion

This study was undertaken primarily to obtain information about the use of motion pictures for projective testing. While the results do not permit valid conclusions to be drawn about the relative effectiveness of films per se, some subjective impressions of the experimenter might appropriately be discussed here.

One of the severest handicaps that a film suffers, when used as in this experiment, lies in the short time which may be given to *S* to view the stimulus scene. (In this instance, two showings of 35 seconds each.) In the conventional TAT administration, *S* may take virtually as much time as he wishes for looking at the card. Furthermore, he has the card before him while he is telling his story. It may well be that this makes an important contribution to the value of a still picture for eliciting projection because, with an essentially unlimited time to study the picture, *S* is provided with an always present source of material from which to elaborate his fantasies.

It is rather difficult to provide as much time in a motion picture as an *S* might effectively use, while retaining the necessary element of ambiguity and adding the presumably important element of realism. A motion picture is realistic partly because it depicts people behaving—acting and interacting. Normally, the more a person is shown behaving, the more the situation is defined, i. e., it loses ambiguity. While motion pictures can be produced showing people behaving without defining the situation, it is not certain, since it is not one's usual experience with motion pictures or with life, that such a procedure does not detract from the realism of the scene. A film loop device which would permit the repeated showing of a single short sequence could be used but, again, it is not certain that this would not impair the realism of the scene inasmuch as people normally do not repeat an identical action over and over. The same difficulty applies to sound, for it would seem that an ambiguous sound heard repeatedly either ceases to be ambiguous, or convinces the subject that it was not intended to be interpretable.

These considerations are subjective and should not preclude further research in the area. They do, however, indicate some of the problems which must be met if a motion picture test similar to the TAT is to be developed.

Conclusions

1. The perception of an ambiguous protagonist as being like oneself does not appear to be as generalized a phenomenon as some of the TAT literature would indicate. Such projection was found to occur but to a very limited extent.

2. The experiment did not yield empirical evidence to determine the probable effectiveness of motion pictures as a medium for projective personality testing. That is, projection was not found to vary with the iconicity of the stimuli. There is subjective evidence, however, to suggest that a film test following the TAT paradigm may not be the most effective use of the motion picture medium for this purpose.

3. With respect to sex and age, this study did not support the hypothesis that *S*s will project more onto protagonists who are physically similar to themselves. Indeed, this study seemed to support the conclusion that such similarity may actually discourage projection.

Summary

A study was undertaken to investigate the probable effectiveness of sound motion pictures as a medium for projective personality testing, and to attempt to define some of the characteristics which should be built into such a test.

Short motion pictures, based upon five TAT cards whose central figures differed by sex and age, were produced in such a way as to incorporate varying degrees of realism.

The projection associated with each of the three variables was measured and compared. Projection was operationally defined as the extent to which a subject's perception of himself agreed with his perception of the protagonist.

Little projection was found to occur. The projection that did occur was associated with the two parental-figure protagonists. No differences in projection associated with the realism variable were found.

Received December 11, 1953.

References

1. Crandall, V. J. Induced frustration and punishment-reward expectancy in TAT stories. *J. consult. Psychol.*, 1951, **15**, 400-404.
2. Eiserer, P. E. The relative effectiveness of motion and still pictures as stimuli for eliciting fantasy stories about adolescent-parent relationships. *Genet. Psychol. Monogr.*, 1949, **39**, 205-278.
3. Eron, L. D. A normative study of the TAT. *Psychol. Monogr.*, 1950, **64**, No. 9 (Whole No. 315).
4. Fulchignoni, E., Sbordoni, Gigliola, & Tarroni, Evelina. Il film e i metodi proiettivi (The film and the projective methods). *Arch. Psicol. Neurol. Psichiat.*, 1950, **11**, 97-169.
5. Gemelli, A. Le film comme méthode projective. *Acta Psychol.*, 1950, **7**, 190-195.
6. Heppell, H. K., & Raimy, V. C. Projective pictures as interview devices. *J. consult. Psychol.*, 1951, **15**, 405-411.
7. Murray, H. A. *Manual for the Thematic Apperception Test.* Cambridge: Harvard Univer. Press, 1943.
8. Riess B. F., Schwartz, E. K., & Cottingham, Alice. An experimental critique of assumptions underlying the Negro version of the TAT. *J. abnorm. soc. Psychol.*, 1950, **45**, 700-709.
9. Rotter, J. B. Thematic apperception tests: suggestions for adminstration and interpretation. *J. Pers.*, 1946, **15**, 70-92.
10. Schwartz, E. K., Riess, B. F., & Cottingham, Alice. Further critical evaluation of the Negro version of the TAT. *J. proj. Tech.*, 1951, **15**, 394-400.

11. Symonds, P. M. Criteria for the selection of pictures for the investigation of adolescent fantasies. *J. abnorm. soc. Psychol.*, 1939, **34**, 271-274.
12. Thompson, C. E. The Thompson modification of the TAT. *Rorschach Res. Exch.*, 1949, **13**, 469-478.
13. Weisskopf, Edith A. A transcendence index as a proposed measure in the TAT. *J. Psychol.*, 1950, **29**, 379-390.
14. Weisskopf, Edith A., & Dunlevy, G. P. Body similarity between subject and central figure in the TAT as an influence on projection. *J. abnorm. soc. Psychol.*, 1952, **47**, 441-445.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

A Note on the Use of a Paper-and-Pencil Form of the MMPI *Hs* Scale for Hospital Use [1, 2, 3]

Sol Charen
Catholic University of America

The use of a single scale of the MMPI, according to Machover and Anderson [1], is often desirable in clinical practice. This is especially true in hospital clinics and outpatient departments where some method of screening patients with neurotic symptoms from those with organic problems is needed.

The 33 items of the *Hs* scale of the MMPI were mimeographed on a single sheet in the same sequence in which they appear in the group form booklet of the MMPI. Subjects were 50 hospitalized tuberculosis patients shortly to be discharged as arrested cases. Their median age was 35 years with the age range from 21 to 69 years.

The method of testing was identical to that used by Machover and Anderson except that the time interval between test and retest was one week. The group-form booklet of the MMPI was used and testing was done in groups. Half the patients were given the paper-and-pencil form of the *Hs* scale and a week later were retested with the booklet. The other 25 were tested in reverse order.

The mean *T* score of the 50 subjects on the experimental form was 56.14 against a mean of 57.87 for the group booklet form of the scale with p equal to .73 for a two-tailed test. Mean *T* scores for all methods of testing ranged from 55.20 to 59.80. The Pearson product-moment correlation between group booklet form of the *Hs* scale and the paper-and-pencil form was .980.

The paper-and-pencil form of the *Hs* scale of the MMPI gave almost identical *T* scores as a retest with the group booklet form. The correlation between the two was .980. The paper-and-pencil form of the *Hs* scale seems, then, worth while for use in hospital clinics and outpatient departments where quick testing and screening of patients is desirable.

Brief Report
Received June 16, 1954.

Reference

1. Machover, S., & Anderson, Helen J. Validity of a paper-and-pencil form of the MMPI Psychopathic Deviate scale. *J. consult. Psychol.*, 1953, **17**, 459-461.

---

[1] From the Veterans Administration Hospital, Baltimore, Maryland. The statements and conclusions reached in this article are those of the author and do not necessarily reflect the opinion or policy of the Veterans Administration.

[2] The author wishes to thank Dr. Ross McLean, Chief of Medicine, VA Hospital, Baltimore, Maryland for his help in making available the patients used in this research.

[3] An extended report of this study may be obtained without charge from Sol Charen, 200 Rhode Island Avenue, N. E., Washington 2, D. C. or for a fee from the American Documentation Institute. To obtain it from the latter source order Document No. 4320 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Transcendence Indices, Extent of Personality Factors in Fantasy Responses, and the Ambiguity of TAT Cards[1]

Douglas T. Kenny
University of British Columbia

This paper is the third in a series concerned with testing the widely accepted assumption of a monotonic relationship between stimulus ambiguity and extent of personality factors in the evoked fantasy. Any study exploring this relationship must concern itself with two primary methodological considerations, namely, the specifications of the meanings assigned to the concepts of both *stimulus ambiguity* and *personality content* and the way in which these terms are to be measured.

Weisskopf [6, 7] has presented evidence showing that pictures with low physical ambiguity elicit less projection than more highly structured ones. Physical "impoverishment" of the stimulus, by incomplete line tracings and varying exposure time, was used as an indicator of stimulus ambiguity. Projection was defined in terms of a transcendence index which is based on the mean number of responses "which go beyond pure description." In an earlier paper it was pointed out by Bijou and Kenny that such findings " . . . may be used to evaluate the proposition under consideration provided one assumes that (*a*) stimulus ambiguity is properly defined in terms of physical properties and, (*b*) there is a direct relationship between amount of fantasy and amount of personality factors introduced" [1, p. 204].

Employing a different set of measurement operations, Kenny and Bijou [2] found that with increasing psychological ambiguity there is an initial increase in the extent of personality factors revealed and then a decrease. In contrast to Weisskopf's procedure, ambiguity was defined in terms of psychological rather than physical properties. Ambiguity was defined in terms of the estimated number of interpretations that judges thought would apply to each TAT picture. In analyzing the significance of personality factors based on TAT stories, Stephenson's [4] *Q*-array technique was utilized rather than unidimensional scales or transcendence indices.

It seems pertinent to relate these two different approaches that have been used in testing the monotonic relationship postulated between stimulus ambiguity and extent of personality factors revealed in fantasy responses. The present research aimed at investigating the relationship between projection, as defined by transcendence indices, and extent of personality factors revealed to TAT pictures, as determined by a *Q*-array technique. It seems worth while to test Weisskopf's assumption that there is a direct relationship between her transcendence indices and the extent of personality factors introduced into TAT protocols. As yet, there is no empirical evidence to indicate that transcendence indices can be used as indicators of the amount of clinically interpretable material elicited by a picture. In addition, an attempt was made to see how transcendence indices varied as a function of physical ambiguity, psychological ambiguity, and stimulus ambiguity involving a combination of both physical and psychological properties.

Method

Ascertaining Stimulus Ambiguity

The three procedures for determining stimulus ambiguity were:

Psychological ambiguity. From the first study [1], 15 TAT cards varying in three degrees of psychological ambiguity were obtained. A group of 51 judges ranked 21 TAT cards in terms of the estimated number of interpretations each card possessed. By this procedure three sets of five cards, varying in three degrees of increasing psychological ambiguity, were selected. Cards 5, 15, 18BM, 11, and 19 were classified as the most ambiguous sets. Cards 4, 7BM, 8BM, 12M, and 13MF made up the intermediate group. Cards 12BG, 1, 2, 9BM, and 17BM comprised the least ambiguous set.

---

[1] The author wishes to express his appreciation to Dr. Edith A. Weisskopf-Joelson for providing him with a transcendence index scoring manual. He is indebted to Professor E. S. W. Belyea for photographing and making the projector slides of the TAT pictures.

Physical ambiguity. In order to induce physical ambiguity, changes in the exposure time of the TAT cards were used. For the present study, the 15 TAT cards were photographed with approximately the same intensity as the standard set. The cards, on individual slides, were projected onto a screen, and the subjects were simply asked to describe the projected picture. The pictures were shown to 80 subjects, using three to eight subjects at one time. All subjects were drawn from introductory psychology classes at the University of British Columbia. Forty male subjects described the pictures under a 5-second exposure time (high physical ambiguity), and another 40 male subjects described the pictures under a 2-minute exposure time (low physical ambiguity).

Combined physical and psychological ambiguity. The 15 TAT cards were classified in terms of both physical and psychological dimensions of ambiguity by making use of Murray's placement of the cards. It will be recalled that Murray divided the cards into two groups of ten, cards numbered 10 and below being more structured than those numbered 11 and above. Murray's placement of the cards was based on the physical characteristics of the cards as well as on diagnostic value. Of the 15 TAT cards, seven are from Murray's first 10 and eight are from his second ten.

Measurement of Personality Factors

As indicated, the aim of this research was to determine the relationship between two measures of fantasy responses to TAT pictures, a Q-array technique and transcendence indices. Transcendence indices for each of the 15 TAT cards were obtained by scoring the descriptions that the subjects gave to the cards under the two exposure times. Weisskopf's manual on the scoring for transcendence indices was employed. Separate transcendence indices were worked out for the different exposure times. For a detailed description of the Q-array scoring method and the administration method of the TAT pictures, the reader is referred to the second paper in this series [2]. Suffice it to say here that the 15 TAT cards were administered individually to 18 male college students. Two well-trained clinical psychologists, utilizing the Q-array method, judged the significance of the personality factors "injected" into the TAT stories. The correlation of the scores between the two clinicians was satisfactory ($r =$ .68).

Results

The most important question likely to arise is: What, if any, is the relationship between the ranking of the TAT cards for revealingness of the personality as determined by the Q-array method and the ranking of the same cards in terms of transcendence indices? The rank-order correlations between personality revealingness and the transcendence indices were .62 and .64 for the 5-second and the 2-minute exposure groups respectively. Statistical significance is beyond the .01 level for the ρ of .64 and at the .05 level for the ρ of .62. The data clearly support Weisskopf's contention that transcendence indices are related to personality revealingness in TAT pictures.

A second question asked is: To what extent does the transcendence index vary as a function of the physical ambiguity of the stimulus? Table 1 presents this finding in terms of the difference between exposure times. The data confirm Weisskopf's findings in showing that subjects reliably increase their transcendence indices under long exposure times.

Table 1

Analysis of Variance of Transcendence Indices

Source of variation	df	Mean square	F
Between exposure times	1	1643.26	44.94**
Subjects in same group	78	36.56	
Total between subjects	79		
Psychological ambiguity level	2	73.86	18.37**
Ambiguity × exposure time	2	17.58	4.37*
Pooled subjects × ambiguity	156	4.02	
Total within subjects	160		
Total	239		

* Significant at 5 per cent level.
** Significant at 1 per cent level.

A third question asked is: Do transcendence indices vary with degree of psychological ambiguity when the transcendence scores are classified under three sets of pictures varying in three degrees of increasing psychological ambiguity? The F for Between Psychological Ambiguity Level, as shown in Table 1, indicates significant differences between the three levels of psychological ambiguity. To determine which sets of cards show the differences, *t* tests were made between the sets using the interaction of pooled subjects × ambiguities as the error estimate. This analysis showed that the intermediate ambiguity set had significantly higher transcendence indices than either the most or least ambiguous sets. The differences were significant at the 1 per cent level of confidence. There were, however, no differences in transcendence indices between the most ambiguous and least ambiguous cards. This finding, combined with the previous finding [2] that the cards in the median ambiguous set reveal significantly greater personality factors than those in the least and most ambiguous sets, shows that those cards which have higher transcendence indices are also the cards which reveal the most about personality.

Data in Table 1 also show a significant interaction between length of exposure time and psychological ambiguity level. Examination of the differences involved in this interaction shows that the intermediate ambiguous set yielded higher transcendence indices than the most ambiguous set under long exposure time, but not under short exposure time; also, the intermediate set showed significantly higher transcendence indices than the least ambiguous set, under either short or long exposure time.

The final question asked is: Are there higher transcendence indices for the TAT cards from the last ten than from the first ten? Analysis of variance results showed that there were no differences in the transcendence indices between these sets of cards. Hence, Murray's hypothesis of a difference in degree of projection between these two sets of cards is not supported by the data. The interaction between this twofold classification of ambiguity and time of exposure was not significant.

Discussion

Two of the most interesting findings which emerge from this study are that the extent of personality material revealed in a picture is correlated with the transcendence index of the picture and that cards from the intermediate psychologically ambiguous set yield higher transcendence indices than either the least or most ambiguous sets. The former finding supports Weisskopf's assumption of a direct relationship between amount of transcendence and amount of personality factors. The latter result may be accounted for in terms of three possible factors which are described in greater detail elsewhere [2]. Briefly, transcendence indices may not have increased with increasing psychological ambiguity because (*a*) anxiety, which may mount with increases in ambiguity, inhibited greater projections; (*b*) the cue patterns in the three sets of cards do not vary as assumed; or (*c*) there may have been qualitative differences among the three sets of cards. Any three of these plausible explanations may account for the obtained results.

The results of the present study, as well as Weisskopf's research, do not conform to the theoretical expectation that increases in physical ambiguity should be accompanied by increases in transcendence indices. In terms of the previous formulation of this problem [2], it may be hypothesized that a decreasing number of cue constellations available for selective reactions produces response tendencies which do not contain much in the way of individual projections.

Contrary toMurray's hypothesis, cards from his more structured series did not yield higher transcendence indices than cards from the less structured series. This finding substantiates the previous conclusion [2] that Murray's most ambiguous set does not elicit more significant personality factors than his least ambiguous set. On the other hand, Weisskopf [5] found that the well-structured series actually elicited higher transcendence indices than the least structured series. This discrepant finding between the two studies may be explained in terms of the differences in the TAT cards used in the two studies.

Summary

This paper describes an experiment relating the concepts of transcendence and extent of personality factors in elicited fantasy. The results are consistent with Weisskopf's assumption that the extent of personality material revealed in a TAT picture is correlated with the transcendence index of the picture. In addition, it was shown that with increasing psychological ambiguity there is an initial increase in transcendence and then a decrease. The theoretical expectation that transcendence indices

should increase with increases in physical ambiguity was not substantiated. Finally, the expectation that TAT cards from Murray's second series should elicit higher transcendence indices than cards from his first series was not upheld.

Received January 30, 1954.

References

1. Bijou, S. W., & Kenny, D. T. The ambiguity values of TAT cards. *J. consult. Psychol.*, 1951, **15**, 203-209.
2. Kenny, D. T., & Bijou, S. W. Ambiguity of pictures and extent of personality factors in fantasy responses. *J. consult. Psychol.*, 1953, **17**, 283-288.
3. Murray, H. A. *Thematic Apperception Test.* Cambridge, Mass.: Harvard Univer. Press, 1943.
4. Stephenson, W. *The study of behavior.* Chicago: Univer. of Chicago Press, 1953.
5. Weisskopf, Edith A. A transcendence index as a proposed measure in the TAT. *J. Psychol.*, 1950, **29**, 379-390.
6. Weisskopf, Edith A. An experimental study of the effect of brightness and ambiguity on projection in the Thematic Apperception Test. *J. Psychol.*, 1950, **29**, 407-416.
7. Weisskopf-Joelson, Edith A., & Lynn, D. B. The effect of variations in ambiguity on projection in the Children's Apperception Test. *J. consult. Psychol.*, 1953, **17**, 67-70.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

A Psychotherapy Case Showing Progress but No Decrease in the Discomfort-Relief Quotient

Edward J. Murray, Frank Auld, Jr., and Alice M. White[1]
Yale University

The purpose of this paper is to demonstrate that a case may show progress even though there is no drop in verbal tension. Verbal tension was measured by the Discomfort-Relief Quotient (DRQ), an objective method of determining the tension expressed in the sentences of a patient during psychotherapy interviews. The DRQ was developed by Dollard and Mowrer [4] to measure tension during social casework interviews. Mowrer [12, 13] and Kauffman and Raimy [10] have gone further with it and have concluded that a tension drop during psychotherapy, as measured by the DRQ, indicates therapeutic progress.

The DRQ was compared with a content-analysis procedure which shows changes that can be interpreted as indicating progress. The method of content analysis was developed by the senior author [14] on the basis of concepts of motivation and conflict derived from learning theory and psychoanalytic theory.

The DRQ and the content analysis were applied to the case of Mrs. B., a psychotherapy case which has received a detailed clinical study by Dollard, Auld, and White [2]. The existence of this qualitative study of the case will enable the interested reader to evaluate better the clinical significance of the quantitative results presented here.

The DRQ Study of the Case

The procedure for scoring the DRQ has been described in detail by Dollard and Mowrer [4]. Briefly, each unit is judged as indicating *discomfort* (drive, want, tension, pain, suffering), *relief* (drive reduction, reward, pleasure, relaxation), or *neither*. The DRQ is the number of discomfort units, divided by the number of discomfort units plus relief units. The resulting quotient is multiplied by 100 to provide a scale ranging from 0 (no tension) to 100 (all tension).

In applying the DRQ to the case of Mrs. B., two graduate students independently scored summaries of all 17 therapy hours. These students were psychologically sophisticated but unfamiliar with the case. They scored the summaries in a random order so that they would not be influenced in their scoring by the "plot" of the case or by preconceptions about the direction of tension change. Their agreement was high ($r = +.92$, $df = 15$, $p < .01$). Figure 1 shows the average DRQ of the two scorers for each of the therapy hours. It can be seen that, although there is some fluctuation in the DRQ from hour to hour, there is no consistent trend throughout the course of therapy. There is neither a gradual decline in ver-

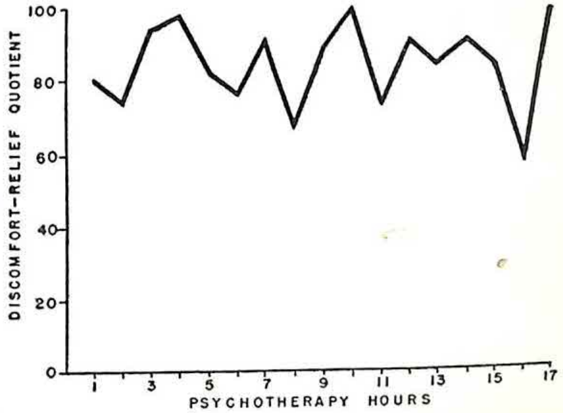


Fig. 1. The Discomfort-Relief Quotient based on summaries of 17 consecutive therapy hours with Mrs. B.

---

[1] Mr. Murray's share of this work was supported by a U. S. Public Health Service research fellowship. Part of the study was also supported by a grant from the U. S. Public Health Service to Professor John Dollard, who sponsored this research. The authors wish to thank Drs. Dollard, George F. Mahl, and Neal E. Miller for a critical reading of the manuscript.

bal tension nor a sudden drop near the end of treatment.

It is possible that the DRQ showed no decline because the summaries did not truly reflect tension expressed in the patient's sentences. In order to explore this possibility, two other psychologically sophisticated people applied the DRQ to the verbatim typescripts of six of the interviews. These scorers were given the pages of the six interviews in random order. The interviews were selected for typing in the following way: five successive blocks of three or four hours were delineated and one hour was selected randomly from each block. Hour 11 was added because it had been previously typed for the Dollard, Auld, and White book. The agreement of the two scorers was high ($r = +.96$, $df = 4$, $p < .01$). The averaged results are shown in Figure 2. Again, no trend is apparent.

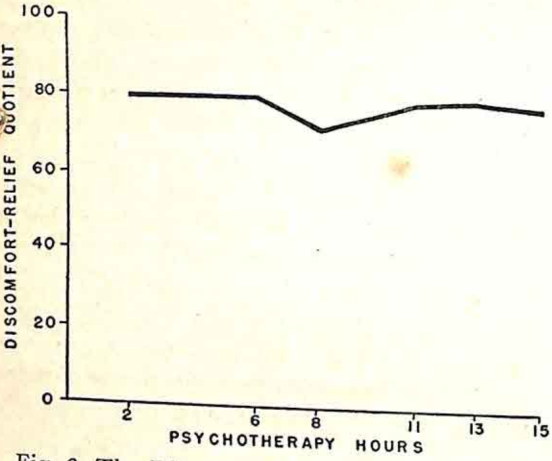


Fig. 2. The Discomfort-Relief Quotient based on verbatim typescripts of six randomly selected therapy hours with Mrs. B.

The Content Analysis

Murray has described his content-analysis method in more detail elsewhere [14]. Scorers in this study put each of the patient's statements into one of several categories. The results are reported for four major categories: (*a*) hostility toward mother and children, (*b*) general hostility toward husband, (*c*) sex conflict related to husband, and (*d*) general anxiety related to no one in particular. These categories include both approach and avoidance tendencies in each area. Thus, hostility includes both expression of hostility and expression of anxiety or guilt about hostility. The sex conflict with the husband includes sexual desire, sexual anxiety, and hostility specifically related to sexual frustration.

The main judge in this study scored the typescripts of six therapy hours in a random order. A psychologically naive scorer was used to check reliability. This person scored two hours with the main judge for training purposes and then scored the remaining four hours independently. Agreement between the two scorers for the four hours was high: for hostility to mother and children, $r = +.997$, $df = 2$, $p < .01$; for general hostility to husband, $r = +.995$, $df = 2$, $p < .01$; for sex conflict with husband, $r = +.993$, $df = 2$, $p < .01$; and for general anxiety, $r = +.956$, $df = 2$, $p < .05$. In order to control for differences in the total number of statements from hour to hour, the raw frequency scores were converted into percentages of the total number of statements in the hour.

The sum of all four categories reveals no striking change throughout therapy, as shown in Figure 3. Whatever slight trend exists can be accounted for almost entirely by the high point on Hour 6. Although the precise shape of this curve differs from the DRQ curve in Figure 2, the over-all impression is the same.

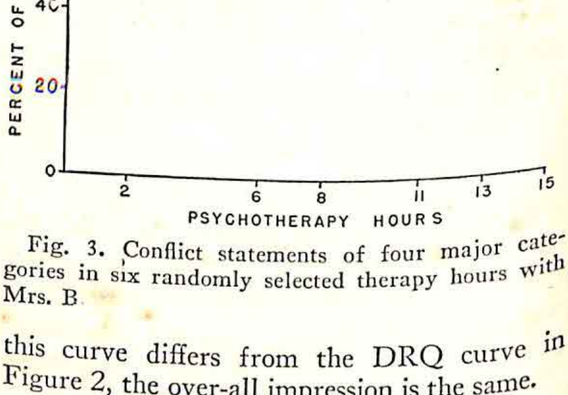


Fig. 3. Conflict statements of four major categories in six randomly selected therapy hours with Mrs. B.

The content analysis shows important changes in the patient's speech when the separate areas of conflict are considered. During the course of therapy, the patient shifted from talking about her conflict with her mother and children to talking about her conflicts with her husband. This is shown in Figure 4, where "hostility conflict related to mother and children" and the sum of the conflict categories relating to the husband (i.e., both general hostility and sex) are plotted.

The nature of the conflicts about the husband also changed during therapy. Figure 5 shows the category of general hostility to the husband and the category of sex conflict with the husband. The general hostility conflict was replaced by the sex conflict during the course of therapy. The general hostility conflict consisted of criticisms of the husband's attitudes toward social life, vacations, etc., and some guilt about these criticisms. The sex conflict included statements about the patient's own frustration, statements blaming her husband, and eventually statements showing some insight into her own inhibitions.

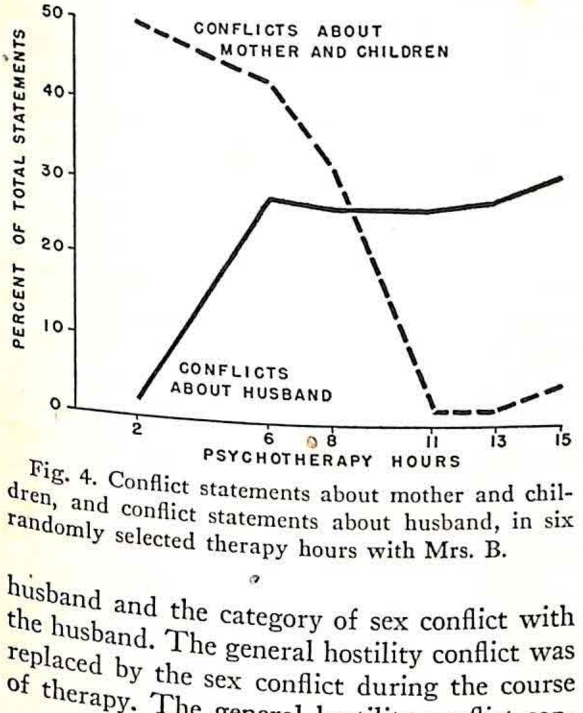


Fig. 4. Conflict statements about mother and children, and conflict statements about husband, in six randomly selected therapy hours with Mrs. B.

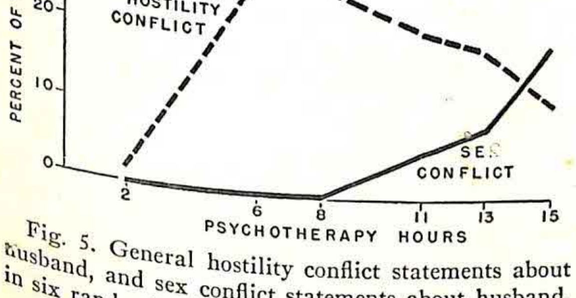


Fig. 5. General hostility conflict statements about husband, and sex conflict statements about husband, in six randomly selected therapy hours with Mrs. B.

The content analysis, then, shows a change during the course of therapy from talking about a hostility conflict about the patient's mother and children, to talking about a hostility conflict about unimportant matters concerning her husband and, finally, to talking about a fundamental sexual conflict about her husband.

Discussion

According to Dollard, Auld, and White [2], Mrs. B. made some progress in therapy, although her case was not a "finished" one. Their account of the case is briefly as follows:

> Mrs. B. began therapy with a defensive "cover story." She complained that her mother had stolen away her children and that she (the patient) was unable to stand up for her rights in the situation. According to Dollard, Auld, and White, this baby-snatching was not a primary problem; indeed, Mrs. B. had very little trouble asserting herself toward her mother or anyone else. The therapist refused to accept the cover story as an adequate explanation. He also insisted that she could not avoid discussing her marital life. Mrs. B. began to complain about her husband. At first she complained about unimportant matters such as his lack of interest in parties. The therapist found it difficult to believe that these things had the overwhelming importance ascribed to them. He noted that the patient had said nothing at all about her sex life. Mrs. B. gradually began to talk about her sex life. She said at first that it was satisfactory; she failed to see any connection between sex and her presenting symptoms of depression. But as she followed the rule of free association she began talking about her sexual frustration and her husband's inadequacy. Later she saw the role her own inhibitions played in the situation. She finally realized that her sexual feelings toward her father had been transferred to her husband along with her horror of committing incest. She had unconsciously rejected her marriage and her responsibilities to her children as well. Toward the end of therapy the patient reported that her husband had spontaneously commented that she "seemed much better in bed — more interested now." Undoubtedly Mrs. B. had not solved all of her problems, but she had gained some insight.

The results of the independent content analysis are in remarkable agreement with the qualitative analysis. The content analysis shows a decrease in the category of "hostility conflict about mother and children," which corresponds to what Dollard, Auld, and White call the "cover story." The content analysis also shows the shift in verbalizations first to a general hostility conflict about the husband and then to a sexual conflict about the husband. The advantage of the content analysis is that it describes precisely and quantitatively therapeutic events which the qualitative analysis describes in an impressionistic way. Moreover, the content analysis has been demonstrated to be reliable. The agreement between the content analysis and the qualitative analysis shows that the content analysis measures meaningful and theoretically relevant things.

The DRQ did not show the decrease which is said to accompany progress. Several other investigations have also found no relationship between the DRQ and therapeutic progress. Hunt [7, 8] and Hunt and Kogan [9] found a correlation of only .20 between DRQ decrement and caseworkers' ratings of movement. The failure of the DRQ to drop in the present case was due to the fact that the patient verbally expressed tension about something or other throughout therapy. The progress in the case was due to an increase in insight which occurred when the verbalizations shifted toward a basic conflict. The patient learned something new about herself which apparently affected her behavior outside of therapy.

An important problem is the relationship of the DRQ to the tension actually experienced by the patient. Tension may have remained at the same level throughout therapy with the patient gradually becoming aware of its origin in her sexual conflict. In this case the DRQ would have accurately reflected the tension experienced by the patient. However, there is reason to doubt such an accurate reflection since the DRQ does not correlate very highly with physiological measures of tension [13]. It is much more likely that Mrs. B. experienced more tension as she began discussing her sexual conflicts than she had while elaborating her cover story. In this connection, it is perhaps significant that the final content category of general, or free-floating, anxiety consistently increased during therapy.

There are many examples in the literature of the need to increase tension in order to get therapeutic results. Freud [6] found it was necessary to break through *la belle indifférence* of the hysteric to achieve therapeutic results. The isolation of emotion from thought in the obsessive-compulsive has proved a stumbling block in therapy; it is only when such a patient *feels* tense that progress is made [5]. Alexander and French [1] have cited many cases in which the frequency of contacts was purposely manipulated to increase tension. The authors believe that ". . . every analysis should be conducted on as high an emotional level as the patient's ego can stand without diminishing its capacity for insight" [1, p. 30]. Miller's theory of conflict [11] provides a rationale for these empirical findings. According to this theory, when a therapeutic tactic enables a patient to approach closer to a feared goal, the absolute level of fear is increased. The result has been called the negative therapeutic effect [3], since the increased fear often brings about a return of symptoms. Therefore, although we would expect an eventual reduction of tension in a successful patient's everyday life, we would not necessarily expect a reduction of tension in the therapeutic situation, particularly in short-term therapy.

The foregoing analysis raises the problem of why the patient stays in therapy in the face of increasing tension. The answer is partly that many patients leave therapy when anxiety increases. In successful cases, the tendency to leave seems to be counterbalanced by several little understood factors. First, there are many rewarding features of the transference relationship [3]. Second, the symptoms and frustrated drives which originally motivated the patient to seek help continue to operate. Third, the solution of minor problems eases the tension in the patient's daily life to some extent and reinforces hope of solving the more important problems.

Summary and Conclusions

1. The purpose of this study was to show that a psychotherapy case in which there was no drop in verbal tension might show progress in other ways. The Discomfort-Relief Quotient was used as a measure of verbal tension,

and a content analysis in terms of motivation and conflict was used to show other verbal changes.

2. Two separate applications of the DRQ to the case of Mrs. B. showed no change in verbal tension during therapy. Each application was demonstrated to be reliable.

3. The content analysis showed that during therapy the patient progressed from a discussion of a "cover story" about a hostility conflict with her mother to a discussion of conflicts related to her husband. Furthermore, she progressed from discussing a general hostility conflict about relatively unimportant matters in relation to her husband to discussing much more fundamental sexual conflicts. The content analysis was shown to be reliable.

4. The results of the content analysis were shown to be consistent with an independent, qualitative clinical analysis.

5. It was pointed out that the DRQ may not reflect the tension experienced by a patient. It was also pointed out that although an eventual tension reduction in a successful patient's everyday life would be expected, a reduction of tension in the therapeutic situation would not necessarily be expected.

6. We conclude that although the DRQ may be a good measure of the tension expressed in a patient's sentences, a study of this one case provides no support for the view that the DRQ is useful for assessing therapeutic progress. In this case, other measures of the content of the patient's speech proved more illuminating than the DRQ.

Received February 8, 1954.

References

1. Alexander, F., & French, T. M. *Psychoanalytic therapy.* New York: Ronald, 1946.
2. Dollard, J., Auld, F., & White, Alice M. *Steps in psychotherapy.* New York: Macmillan, 1953.
3. Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
4. Dollard, J., & Mowrer, O. H. A method of measuring tension in written documents. *J. abnorm. soc. Psychol.,* 1947, **42,** 3-32.
5. Freud, S. *Collected papers.* Vol. II. London: Hogarth, 1924.
6. Freud, S. *Collected papers.* Vol. IV. London: Hogarth, 1925.
7. Hunt, J. McV. Measuring the effects of social casework. *Trans. N. Y. Acad. Sci., Ser. II,* 1947, **9,** 78-88.
8. Hunt, J. McV. A social agency as a setting for research — the Institute of Welfare Research. *J. consult. Psychol.,* 1949, **13,** 69-81.
9. Hunt, J. McV., & Kogan, L. S. *Measuring results in social casework.* New York: Family Service Association of America, 1950.
10. Kauffman, P. E., & Raimy, V. C. Two methods of assessing therapeutic progress. *J. abnorm. soc. Psychol.,* 1949, **44,** 379-385.
11. Miller, N. E. Comments on theoretical models, illustrated by the development of a theory of conflict behavior. *J. Pers.,* 1951, **20,** 82-100.
12. Mowrer, O. H., Hunt, J. McV., & Kogan, L. S. Further studies utilizing the Discomfort-Relief Quotient. In O. H. Mowrer (Ed.), *Psychotherapy: theory and research.* New York: Ronald, 1953. Pp. 257-295.
13. Mowrer, O. H., Light, B. H., Luria, A., & Zeleny, M. P. Tension changes during psychotherapy with special reference to resistance. In O. H. Mowrer (Ed.), *Psychotherapy: theory and research.* New York: Ronald, 1953. Pp. 546-640.
14. Murray, E. J. A case study in a behavioral analysis of psychotherapy. *J. abnorm. soc. Psychol.,* 1954, **49,** 305-310.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

The Effect of Childhood Diseases on IQ Variability[1]

Lillian Wagenheim
The University of Southern California

Two previous studies by the writer showed a relationship between the early contraction of measles, chicken pox, mumps, and German measles and later school achievement. This relationship was found among boys who had contracted these diseases under age three and most significantly in those at age two.

In the present study, the records of 493 boys were used. The data obtained included IQ's, achievement grade level in reading and arithmetic, and the ages of contraction of the following childhood diseases: measles, chicken pox, German measles, mumps, whooping cough, diphtheria, scarlet fever, tonsillitis, pneumonia. Cases were considered variable if two IQ scores resulting from tests administered during different years showed a difference of 10 or more points.

The contraction of whooping cough, tonsillitis, chicken pox, and measles under age five was related at the 1% or 3% level to the variability in IQ. The contraction of mumps, German measles, and pneumonia under age five did not show a significant relationship to IQ variability.

In order to discover the effect of childhood diseases in general on IQ variability, the cases were regrouped for the earliest contraction of any disease.

Boys who had contracted any disease under age five were more frequently found in the variable group than in the nonvariable group. Chi square was 11.65 and significant beyond the 1% level of confidence.

An especially interesting and relevant finding occurred when the records of children of different school achievement levels were grouped separately. The cases had been divided on the basis of achievement in reading and arithmetic for a previous study. A child who functioned at least .7 of a grade below his chronological age level was considered an underachiever.

Results significant at the 1% or 2% level were found in the separate achievement groups at different age levels. The children who were underachievers in both reading and arithmetic had the lowest median IQ. They also showed more variability in IQ when childhood diseases had been contracted at ages two through four. Children who were underachievers in *either* reading *or* arithmetic were average in IQ, and their intelligence was more variable when diseases had been contracted at ages one through three. Those who were achieving in both reading and arithmetic were superior in median IQ and showed more variability in intelligence when the diseases were contracted at ages zero through two years.

It seems premature to attempt to answer the question as to whether the disturbance has a direct relationship to encephalomyelitis following childhood diseases or whether the trauma to the nervous system interferes with the psychosocial (ego) development of the child. Apparently there is a hierarchy of ages at which the diseases are most deleterious and which appears to coincide with the emergence of language ideation at different levels of intelligence.

Brief Report
Received June 21, 1954.

---

[1] An extended report of this study may be obtained without charge from Lillian Wagenheim, University of Southern California, Los Angeles 7, California or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4319 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

The Relationship Between Motor and Cognitive Inhibition

Julian Meltzoff[1]
VA Regional Office, Brooklyn, N. Y.
and Murray Levine
Philadelphia VA Mental Hygiene Clinic

In psychoanalytic theory, the gradual shift from the primary to the secondary process involves the delay and control of impulse expression and the substitution of symbolization, fantasy, and planning for immediate response on a behavioral level. Rapaport's [5] psychoanalytic theory of thinking points to the inhibition of action designed to gratify immediate needs as a primary condition for the development of fantasy. Young children are characteristically lacking in the ability to inhibit motor activity and tend to verbalize whatever thoughts come to mind. Piaget [3] and Luria [1] have illustrated the use of speech and symbolic circuition for the control of behavior. This is paralleled in the Rorschach test by a gradual emergence of human movement (M) responses with maturation [9]. Certain psychotic groups also display faulty inhibition of thoughts and motor behavior, and are similarly lacking in the ability to produce M.

Rorschach [6] implied a triadic relationship between motor activity, motion perception, and the cognitive processes. Experiments by Singer, Meltzoff *et al.* [2, 7, 8], testing hypotheses deduced from Werner and Wapner's [11] sensory-tonic theory, have demonstrated one aspect of that triadic relationship, namely the vicarious relationship between motor inhibition and illusory motion perception. The purpose of the present investigation is to study the linkage between motor inhibition and the inhibition of a cognitive process.

The Experiment

Hypothesis. It is hypothesized that there is a direct relationship between motor inhibition and the inhibition of an association process. More specifically, it was predicted that *S*s with greater ability to voluntarily inhibit motor activity would be able to inhibit learned associations and produce new ones quicker than those less adept at inhibiting motor activity.

Procedure. Eighty volunteer male and female *S*s were selected haphazardly from a university population. They were requested to memorize a verbally presented list of ten pairs of words, to a criterion of one perfect reproduction. Instructions were those used in the Wechsler Memory Scale [10]. The list of pairs was composed of stimulus words and the most common associations to them, as reported by Rapaport [4]. Association time was recorded to the nearest half second.[2] After the list was memorized, *S*s were instructed to respond, upon presentation of the stimulus word, with any word *other than* the learned paired associate. Responses were timed in the same manner as previously. The *S*s were then instructed to write the phrase "New Jersey Chamber of Commerce," as slowly as possible, within the boundaries of parallel lines drawn one-half inch apart. They were reminded at 45-second intervals to write as slowly as possible.

The motor inhibition task of enforced slow writing had been used successfully in prior experiments [2]. As far as the cognitive inhibi-

---

[1] The authors are greatly indebted to VA trainees Louis Pietroforte, Norman Slamecka, and Norris Weinberg for their assistance in this experiment, and to Jerome and Dorothy Singer for their stimulating criticisms and suggestions.

[2] In the experimental procedure, the dependent variable was measured prior to the independent variable, so that any errors in timing were not biased errors.

tion task is concerned, it was reasoned that the production of new associations to words with already memorized associates would involve two processes: (*a*) inhibiting the learned associate and (*b*) finding and reporting a different association. We shall refer to the latter process as *associative facility*. There was no apparent reason to believe it to be correlated with the ability to inhibit motor activity. The assumption was therefore made that associative facility would vary randomly among the experimental population, and that any obtained differences would be primarily due to systematic variation in the ability to inhibit associations. A partial test of this assumption was subsequently carried out, and is reported in the discussion of results. The *S*s' introspections about the cognitive inhibition task were solicited, and they uniformly reported either a conscious attempt to suppress an image of the learned associate, or, in the absence of an image, an awareness that there was a word to be avoided before an association could be produced.

Treatment of the data. The population of 80 *S*s was divided into long and short motor inhibition time (MIT) groups at the median of the distribution of obtained inhibition times. It is of interest to note the extreme differences among *S*s in their capacity to inhibit motor activity. Time for writing the experimental phrase varied all the way from 20 seconds to 44 minutes, 36 seconds. The significance of the difference between the means of these two groups of 40 *S*s on cognitive inhibition time (CIT) was calculated by Fisher's *t*. Analysis of covariance was carried out to take into consideration differences between the reaction times of the groups on the initial criterion trial.

Results

As can be seen in Table 1, the mean CIT of the poor motor inhibitors (i.e., those with short MIT) is appreciably greater than that of the good motor inhibitors (i.e., those with long MIT). This difference is statistically significant at $p < .005$, with a t of 3.38. Analysis of covariance reveals that this difference cannot be accounted for by the initial differences in reaction time that exist between the two groups. The obtained F value is significant at $p < .005$.

There was relatively little overlapping between the two groups. Seventy-seven per cent of *S*s who were least able to inhibit motor activity took longer to inhibit associations than the mean time of the better motor inhibitors. On the other hand, only 12 per cent of the better motor inhibitors took longer than the mean time of the short MIT group to inhibit associations.

Table 1

Mean Reaction Time and Cognitive Inhibition Time of Long and Short Motor Inhibition Time Groups

Measure	Short MIT group	Long MIT group
Mean reaction time, Paired associates*	1.0	0.9
Mean CIT*	2.9	2.0

* Times are group means of individual *S*s' average reaction times for ten responses.

Table 2

Analysis of Covariance of Reaction Time and Cognitive Inhibition Time of Long and Short Motor Inhibition Time Groups

Source	*df*	Adjusted mean square	*F*
Between groups	1	1278.680	9.94
Within groups	77	128.455	

Discussion

In view of the fact that the experimental sample was divided at the median, instead of representing extremes of the population, the clear difference between the groups lends strong confirmation to the hypothesis. The ability voluntarily to inhibit motor activity apparently goes hand in hand with the ability to inhibit associations as measured by the experimental task. This finding, however, is dependent upon the correctness of the assumption underlying the choice of the particular task selected to measure thinking inhibition. The primary issue is the relegation of associative facility to a secondary role while considering inhibition as the primary determinant of association times in the experimental task.

In order to provide an experimental check on this reasoning, the cognitive inhibition task was repeated with 30 new *S*s from the same general population. These *S*s were also required to give a single association to each of ten different words taken from the same source as the original pairs. The response time was taken as a measure of associative facility. A random half of the *S*s had the pairs first, while the remainder began with the single-word associations. Division of the population into fast and slow CIT groups reveals a mean association time of 1.3 seconds for the rapid thought inhibitors and a mean of 1.6 seconds for the slow group. This difference is in the inconclusive range, with a t of 1.72 ($p = .10$). There is a positive, but not statistically significant, correlation (rho $= .32$) between the two variables.

These results suggest that associative facility is perhaps more of a factor than was originally assumed, but nevertheless a secondary one which could not reasonably account for very much of the highly significant difference between the groups in the original experiment. It is quite plausible that the ability to produce new associations rapidly is *one* of the means by which people are able to accomplish the work of inhibiting thoughts. Those who have this ability are to this extent better equipped to inhibit. There are probably other abilities as well that facilitate inhibition.

The findings of this investigation support the speculation and evidence from diverse sources relevant to the development of the secondary process cited in the introduction. The results also bear, at least indirectly, upon our understanding of Rorschach *M* responses. Werner's sensory-tonic theory suggests a vicarious relationship between perceptual and motor activity and enables one to predict that a decrease in motor activity will be accompanied by an increase in the perception of illusory motion. Experiments by Singer, Meltzoff *et al.* [2, 7, 8] confirm hypotheses deduced from this theory and show that the voluntary inhibition of motor activity results in an increase in *M* production. They also have demonstrated that those who are more productive of *M* responses are characteristically better able to inhibit motor activity. The present study adds the additional information that those who are more capable of inhibiting motor activity are also better able to inhibit established associations and produce new ones. It does not necessarily follow that *M* is a measure of the capacity to inhibit associations as well as motor activity, but these accumulated data would lead us to hypothesize that such a relationship exists.

It is of interest to report the application of these experimental techniques to the study of the psychopathology of an individual clinical case. The patient was an intelligent manic whose behavior was characterized by motor and verbal hyperactivity and flights of ideas. He was examined while under treatment with the drug Tolserol, a muscular relaxant. He displayed no atypical motor behavior during the course of the examination. He gave neither *M* nor color responses on the Rorschach, could inhibit his motor activity for only 29 seconds (the mean of the total experimental population is 5 minutes, 48 seconds),[3] and was very poor at inhibiting associations with a mean time of 6.4 seconds (the mean of the total experimental population was 2.4 seconds).[3] It was concluded that despite the superficial effects of Tolserol, the patient was basically unable to inhibit either motor activity or thoughts. About two weeks later, medication was discontinued, and he had a violent episode marked by uninhibited verbalizations and extreme hyperactivity necessitating his being placed in seclusion.

Summary and Conclusions

This study was designed to test the hypothesis of a direct relationship between motor inhibition and the inhibition of a cognitive process. Eighty *S*s were divided at the median into two groups on the basis of motor inhibition time, as measured by an enforced slow writing task. The difference between these two groups in cognitive inhibition time, as measured by *S*s' speed in inhibiting learned word association pairs and producing new associations to the stimulus words, was found to be statistically significant in the predicted direction.

It is concluded that *S*s with greater ability to voluntarily inhibit motor activity are able to inhibit learned associations and produce new ones more quickly than those less adept at inhibiting motor activity.

Received January 18, 1954.

---

[3] These means are merely illustrative. They are *not* norms for clinical use.

References

1. Luria, A. R. *The nature of human conflicts.* New York: Liveright, 1932.
2. Meltzoff, J., Singer, J. L., & Korchin, S. J., Motor inhibition and Rorschach movement responses: a test of the sensory-tonic theory. *J. Pers.,* 1953, **21**, 400-410.
3. Piaget, J. *The language and thought of the child.* New York: Harcourt, Brace, 1932.
4. Rapaport, D. *Diagnostic psychological testing.* Vol. 2. Chicago: Year Book Publishers, 1946.
5. Rapaport, D. *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951.
6. Rorschach, H. *Psychodiagnostics.* New York: Grune & Stratton, 1942.
7. Singer, J. L., Meltzoff, J., & Goldman, G. D. Rorschach movement responses following motor inhibition and hyperactivity. *J. consult. Psychol.,* 1952, **16**, 359-364.
8. Singer, J. L. & Spohn, H. E. Some behavioral correlates of Rorschach's experience-type. *J. consult Psychol.,* 1954, **18**, 1-9.
9. Thetford, W., Molish, H. B., & Beck, S. J. Developmental aspects of personality structure in normal children. *J. proj. Tech.,* 1951, **15**, 58-78.
10. Wechsler, D. A standardized memory scale for clinical use. *J. Psychol.,* 1945, **19**, 87-95.
11. Werner, H., & Wapner, S. Sensory-tonic field theory of perception. *J. Pers.,* 1949, **18**, 88-107.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Classification of "Suicidal" Patients[1]

Albert Rosen, William M. Hales,[2] and Werner Simon
VA Hospital, Minneapolis, Minnesota and University of Minnesota

The term "suicidal" has been used to describe patients demonstrating heterogeneous kinds of behavior, such as (*a*) suicide thoughts, (*b*) suicide threats, (*c*) suicide attempts, and (*d*) severe depression. Some investigators have grouped together psychological test data from such individuals, as well as from those who later committed suicide, in an attempt to develop and evaluate psychometric devices for the detection of "suicidal" patients [2, 3, 4, 5, 6]. It is important to consider whether or not such grouping of patients under the class term "suicidal" is too broad for effective use of psychological tests in the study and detection of the patient who is apt to attempt suicide.

Only one published study is relevant to the problem of differentiating "suicidal" subgroups by means of psychological tests. Farberow [1] analyzed MMPI records and other tests of three groups of patients in a California veterans psychiatric hospital. Each group contained 32 male patients as follows: (*a*) patients who made a suicide attempt within a relatively short period before admission, (*b*) patients who made suicide threats, and (*c*) patients selected at random, with no history of suicide attempts, threats, or thoughts (Control Group). Tests were administered to each patient in the first group *after* the attempt was made. The most striking finding in the MMPI data was the marked discrepancy in elevation of mean MMPI scale scores between the Suicide Attempt and Suicide Threat Groups. The scores for the Suicide Threat Group indicated a much greater severity of disturbance. Moreover, the mean scores for the Suicide Attempt and the Control Groups were not significantly different from each other.

Farberow's results from the use of other tests also emphasized that patients who have made a suicide attempt are considerably dissimilar from those expressing suicide threats. The purpose of this paper is to report an investigation conducted in the Psychiatry Section of the Minneapolis Veterans Administration Hospital. This study consisted essentially of a replication of that portion of Farberow's research employing the MMPI in the evaluation of "suicidal" patients.

Selection and Description of Subjects

From the summaries of all discharged male patients, two groups were selected, as follows: (*a*) patients who had made a suicide attempt shortly before admission to the hospital (Suicide Attempt Group), and (*b*) patients who expressed suicide thoughts (Suicide Thought Group). Although finer grouping of the cases was desired, it was not possible from the available file data. The MMPI for each of these patients was then obtained, but was included in the study only if it had been administered within 15 days after admission and prior to electroshock or insulin therapy. A few additional cases were acquired from more recent discharge summaries in order to obtain samples of 50 Suicide Attempt cases and of 100 Suicide Thought cases.

A third sample consisting of a cross section of the psychiatric patient population was included in the study for comparison with the Suicide Attempt and the Suicide Thought Groups. In a previous unpublished study, a sample of 250 psychiatric patients-in-general had been selected. This group consisted essentially of every third patient admitted during a two-year period who had taken the MMPI. For the purposes of the present study, 39 cases were removed from this patient-in-general group because their discharge summaries indicated one or more attempts at, or thoughts of, suicide. These 211 cases will be referred to as the Control Group.

---

[1] From the psychiatric service, VA Hospital, Minneapolis, Minnesota, and the Divisions of Psychiatry and Clinical Psychology of the University of Minnesota.

[2] W. M. Hales is now at the Area Medical Office, Veterans Adminstration, Fort Snelling, St. Paul, Minnesota.

For each group, data were obtained on diagnosis, age, education, marital status, and intelligence. Only summary statistics will be presented because the major interest of this investigation is in the comparison of personality test scores. Primary diagnoses of the Suicide Attempt, the Suicide Thought, and the Control Groups are listed in Table 1. The groups do not differ significantly in broad diagnostic categories. However, among patients classified as neurotic, those who have made suicide attempts or who have engaged in suicide thinking are much more frequently diagnosed as depressive reaction than are patients-in-general. It is quite possible that the diagnosis of depressive reaction is sometimes influenced by the evidence of suicidal behavior or mentation.

Biographical data for the Suicide Attempt, Suicide Thought, and Control Groups, respectively, are as follows: median years of age—32, 32, 32; median years of education—11, 12, 10; percentage of group who are married or widowed—68, 52, 58; percentage who are divorced or separated—8, 18, 12; percentage in group below 110 IQ as roughly estimated from the Shipley-Hartford test—26, 38, 44.

Table 1

Primary Diagnoses of Patients in the Suicide Attempt, Suicide Thought, and Control Groups

Diagnosis	Per cent of Attempt Group ($N = 50$)	Per cent of Thought Group ($N = 100$)	Per cent of Control Group ($N = 211$)
Depressive reaction	34	22	4
Anxiety reaction	16	24	28
Other neurotic	6	7	26
Total neurotic	56	53	58
Paranoid schizophrenia	16	14	10
Involutional melancholia	0	7	4
Other psychotic	4	8	8
Total psychotic	20	29	22
Behavior disorder	18	15	15
Organic	4	3	3
No diagnosis	2	0	2
Total sample	100	100	100

The median number of days between hospital admission and MMPI testing for the Suicide Attempt, Suicide Thought, and Control Groups was 3, 4, and 3, respectively.

In the Suicide Attempt Group, the number of patients making each of the various modes of attempt is as follows: Sedation—18; slash-

Table 2

MMPI Scale Raw Scores for Suicide Thought, Suicide Attempt, and Control Groups

Scale	Mean raw score			Variance raw score			*F* ratio
	Thought	Attempt	Control	Thought	Attempt	Control	
L	3.63	3.28	3.76	6.37	6.68	4.36	.88
F	9.63	8.68	7.40	40.83	39.62	27.72	5.28**
K	12.15	12.64	13.73	25.41	36.87	30.04	3.07*
Hs	14.37	11.14	13.76	48.73	75.04	49.77	3.63*
Hs + .5 *K*	20.68	17.74	20.88	41.86	45.47	42.47	4.77**
D	33.08	28.14	28.60	40.71	44.28	49.93	16.16**
Hy	29.76	26.66	28.47	41.76	53.70	43.58	3.64*
Pd	23.28	23.04	20.92	36.38	50.44	37.54	5.83**
Pd + .4 *K*	28.24	28.26	26.40	32.26	32.27	32.60	4.58*
Mf	26.93	26.08	25.33	28.51	29.83	29.60	2.99
Pa	14.40	13.46	12.31	19.90	24.89	14.73	8.66**
Pt	26.97	22.70	21.38	97.99	151.89	115.74	9.11**
Pt + 1.0 *K*	39.10	35.34	35.12	57.89	70.54	60.13	9.15**
Sc	26.74	22.40	20.22	180.43	222.80	150.78	8.47**
Sc + 1.0 *K*	38.87	35.04	33.95	128.63	117.92	92.31	7.73**
Ma	18.20	19.58	17.76	23.80	33.56	28.33	2.41
Ma + .2 *K*	20.58	22.14	20.47	19.50	25.16	23.15	2.57
Si	37.82	32.00	32.10	118.09	134.48	120.18	9.72**

* Significant at the .05 level.
** Significant at the .01 level.

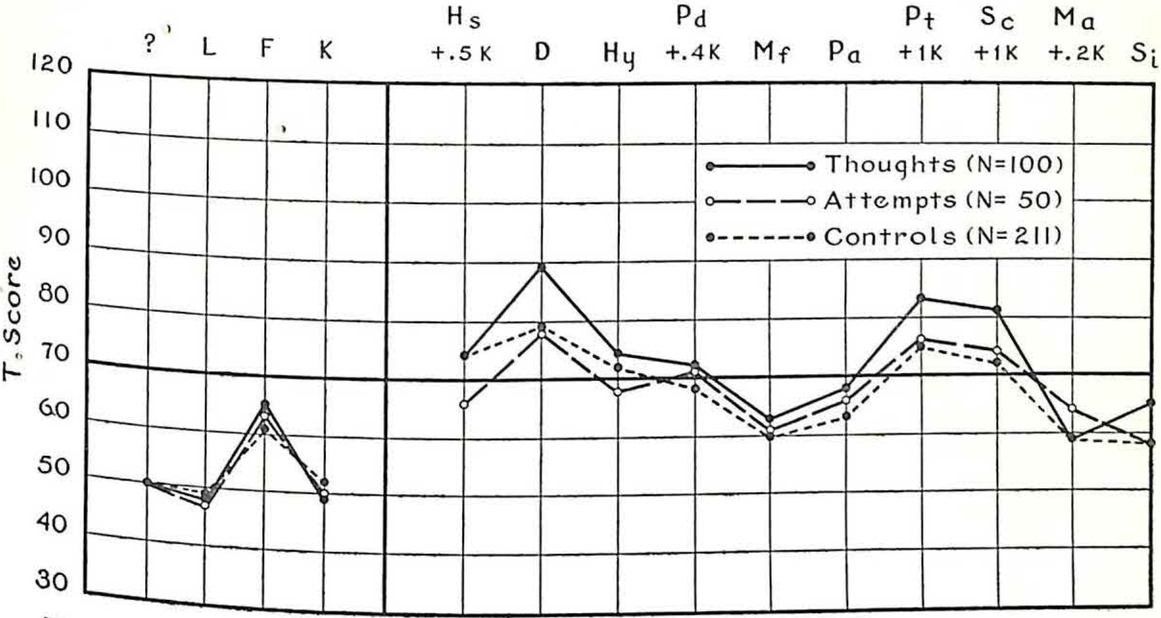


Fig. 1. Mean MMPI T-score profiles for Suicide Thought, Suicide Attempt, and Control Groups.

ing wrists—8; gas or carbon monoxide—7; shooting — 4; hanging or strangulation — 3; jumping from bridge—2; jumping in path of auto—2. For 6 cases the method was not indicated. For 2 cases listed only as "slashing," one also used strangulation, and another gas.

MMPI Results

All calculations were made with raw scores, and the raw score mean and variance of each MMPI scale for the Suicide Attempt, Suicide Thought, and Control Groups are presented in Table 2 for reference purposes. The Cannot Say score was zero in most cases, and is therefore not considered further. Mean T scores based on the raw scores are shown in Figure 1. In order to determine the significance of the differences between means, an analysis of variance test was made for each scale. The Bartlett test indicated no significant difference in the variances for any scale. The F ratio for each analysis of variance is listed in the last column of Table 2, from which it can be seen that almost every scale differentiates in some manner between the groups. For each scale with a significant F ratio the significance of the difference between the means of each pair of the three groups was computed by the t test. Only differences found to be significant at or beyond the .05 level are listed in Table 3.

Table 3

t Ratios for Significant Differences in Mean MMPI Scores between Suicide Thought, Suicide Attempt, and Control Groups

Scale	Thought>Attempt	Thought>Control	Attempt>Control	Control>Attempt	Control>Thought
F		3.22			
K					2.38
Hs	2.62			2.34	
$Hs + .5\ K$	2.59			3.04	
D	4.16	5.38			
Hy	2.67				
Pd		3.10	2.15		
$Pd + .4\ K$		2.65	2.07		
Pa		4.09			
Pt	2.01	4.26			
$Pt + 1.0\ K$	2.77	4.18			
Sc		4.11			
$Sc + 1.0\ K$		3.92			
Si	2.69	4.26			

Note.—t ratios of 1.96, 2.58, and 3.29 are significant at the .05, .01, and .001 levels, respectively.

On every scale in Table 3, except for K, the Suicide Thought Group is significantly higher in mean score than either the Suicide Attempt or the Control Group. The Suicide Attempt Group is significantly higher than the Control Group on the Pd and $Pd + .4\ K$ scales, and significantly lower on Hs and $Hs + .5\ K$. These findings indicate that, as a group, patients with thoughts of suicide are much more severely disturbed than either patients who have made a suicide attempt or patients-in-general, and that patients in the latter two categories are similar to each other. One might hypothesize that for many individuals, a reduction in the degree of psychiatric disturbance follows a suicide attempt. However, this hypothesis can be evaluated only if one has test data for such individuals both before and after a suicide attempt.

The results of this study are quite consistent with those of Farberow [1]. Both studies indicate that patients who have attempted suicide are markedly different from those who threaten or contemplate suicide. In any investigation of the "suicidal" patient, it appears probable that the most meaningful results would be obtained from an analysis of groups which are as finely delineated as possible.

Summary

Psychological test data from patients demonstrating various kinds of "suicidal" behavior are usually grouped together by investigators who are attempting to develop suicide detection devices. Farberow has shown that there are marked differences between groups of patients who have made suicide attempts and those who have threatened suicide. The present paper reports a replication of a portion of Farberow's research. The MMPI records of three groups of male, veteran psychiatric patients were analyzed. The groups consisted of (*a*) patients who made a suicide attempt shortly before admission (Suicide Attempt Group), (*b*) patients who expressed thoughts of suicide (Suicide Thought Group), and (*c*) patients selected at random, with no evidence of suicide attempts, threats, or thoughts (Control Group). The two studies closely agree in indicating that groups of patients who express suicide mentation and threats are much more severely disturbed than either patients who have made a suicide attempt or patients-in-general, and that patients in the latter two categories are similar to each other. It may be concluded that in any investigation of suicide, most meaningful results might be obtained if data for individuals with various kinds of "suicidal" behavior are differentially classified.

Received December 7, 1953.

References

1. Farberow, N. L. Personality patterns of suicidal mental hospital patients. *Genet. Psychol. Monogr.*, 1950, **42**, 3-79.
2. Fisher, S. The value of the Rorschach for detecting suicidal trends. *J. proj. Tech.*, 1951, **15**, 250-254.
3. Hertz, Marguerite R. Suicidal configurations in Rorschach records. *Rorsch. Res. Exch. & J. proj. Tech.*, 1948 **12**, 3-58.
4. Hertz, Marguerite R. Further study of "suicidal" configurations in Rorschach records. *Rorsch. Res. Exch. & J. proj. Tech.*, 1949, **13**, 44-73.
5. Simon, W., & Hales, W. M. Note on a suicide key in the Minnesota Multiphasic Personality Inventory. *Amer. J. Psychiat.*, 1949, **106**, 222-223.
6. White, Mary A., & Schreiber, Hanna. Diagnosing "suicidal risks" on the Rorschach. *Psychiat. Quart. Suppl.*, 1952, **26**, 161-189.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Duration of Illness as a Prognostic Indicator in Mental Illness

C. H. Swensen, Jr.[1] and G. R. Pascal[1]
University of Tennessee

There is unanimous agreement among professional workers in the field of psychiatry that the sooner treatment begins after the development of mental illness, the greater the chances for recovery. As Kalinowsky and Hoch put it, "The most important factor in evaluating prognosis today is the duration of illness prior to treatment; complete agreement exists among all workers that there is an inverse relationship between the duration of illness and recovery" [1, p. 84].

It was the purpose of this study to evaluate the prognostic significance of duration of illness by studying its relationship to outcome of illness after several other variables, also significantly related to outcome, had been controlled.

Method

The data used in this study were obtained from 486 cases admitted to Western Psychiatric Institute and Clinic (WPIC), University of Pittsburgh, between November 1942 and July 1950. These cases included all persons admitted who met the following criteria: age between 15 and 55, no known brain damage, no mental deficiency, no serious physical illness, and no previous attacks of mental illness. This sample is more fully described in a previous paper [2].

The patients were judged to be either "improved" or "unimproved" as of one year following their discharge from WPIC. Patients were judged on the basis of social workers' or psychiatrists' reports. They were considered improved if they were out and working on a level comparable to that on which they had worked prior to becoming ill, and they were judged unimproved if still in a hospital, or if they were at home being cared for.

Duration was measured in months; the time that elapsed between the first time the patient or relatives sought medical help for the patient and the time he was admitted to the hospital for treatment was considered the duration. It is admitted that this allows for a considerable period of time for the patient to have been ill before it was realized by his relatives, particularly in the case of those whose illnesses developed insidiously, but this criterion was used because it was felt to be the most reliable for the purpose of this study.

The relationship between outcome of illness and 21 variables was determined. These 21 variables were: sex, age, marital status, nationality, occupation, economic level, body type, education, intellectual status, affective expression, orientation, direction of aggression, diagnosis, age at onset of illness, duration of illness, length of hospitalization, status upon discharge, precipitating stress, treatment, religion, and type of onset. It was found that 11 of these variables were related to outcome at the .01

Table 1

Summary Statistics of the Total Groups and Controlled Sample Group on Duration of Illness

Statistic	Total group	Controlled sample group
N	486	180
Mean months		
Improved	21.70	34.08
Unimproved	38.59	23.18
Standard deviation		
Improved	22.74	37.12
Unimproved	27.90	20.59
t†	7.31**	2.45*
r_{bis}‡	.38**	—.14

† Between improved and unimproved within group.
‡ Between duration and outcome.
* Significant at .05 level.
** Significant at .01 level.

---

[1] At the University of Pittsburgh when this research was done.

level of confidence. These 11 were: marital status, affective expression, orientation, direction of aggression, diagnosis, type of onset, duration of illness, hospital stay, status on discharge, precipitating stress, and treatment. The relationship between these variables and outcome is discussed more fully in a previous paper [2].

The correlation between duration of illness and outcome for this total sample of 486 was .38. This is presented in Table 1. It will be noted that the unimproved group had a mean duration of illness of 38.59 months, or 17 months more than the mean for the improved group.

The relationships between duration of illness and the other variables that were significantly related to outcome of illness are contained in Table 2. Duration was related to five of the variables (affective expression, direction of aggression, type of onset, precipitating stress, and treatment) at the .01 level. A long duration of illness tended to be associated with an inappropriate affect or an absence of affect, with aggressive impulses directed toward the environment, with the insidious onset, with relatively little precipitating stress, and the patients with a long duration tended to be treated more often with shock therapy than with other types of therapy.

Table 2

Relationship between Duration of Illness and Other Significant Variables

Variable	Chi square	p
Affective expression	26.9	.01
Orientation	4.2	.05
Direction of aggression	24.3	.01
Diagnosis	11.3	.05
Type of onset	79.8	.01
Precipitating stress	19.3	.01
Treatment	23.5	.01
Marital status	5.2	.30

After the relationships between these variables and outcome had been determined, the total group was divided into two groups, an improved group and an unimproved group. These groups were then matched for all of the significant variables except duration of illness. This matching was accomplished by selectively dropping cases from the two groups until there were no significant differences between the improved and unimproved groups on any of the variables other than duration that were significantly related to outcome of illness. There were 180 cases remaining, 90 improved and 90 unimproved, when this matching had been completed. The differences between the controlled sample groups on the significant variables are presented in Table 3. A check revealed that there were no differences between the groups on any of the variables not significantly related to outcome. After this matching was completed, the relationship between duration of illness and outcome of illness for the controlled groups was determined.

Table 3

Differences between the Controlled Sample Groups in the Significant Variables

Variable	Statistic*
Diagnosis	chi square = .00
Marital status	chi square = .10
Orientation	chi square = .51
Treatment	chi square = .00
Affect	$t = .64$
Direction of aggression	$t = .20$
Type of onset	$t = .89$
Hospital stay	$t = .08$
Precipitating stress	$t = .82$

* All below .10 level of confidence.

Results and Discussion

A comparison of the controlled sample groups revealed that the unimproved group had a lower mean duration of illness than the improved group, and that this difference was significant at the .05 level. The biserial correlation coefficient, which had been .38 for the total group of 486, was reduced to —.14. (These results are contained in Table 1, which was presented earlier.) These figures tend to give the impression that under the controlling procedure used in this study, a longer duration is a better sign than a short duration. However, a comparison of the frequencies having various durations for the improved and unimproved groups, as presented in Table 4, reveals that there is little difference between the unimproved and improved groups on number of patients falling into the various categories of duration. In the process of eliminating the effect of the other variables, the improved group was left with several patients with extremely

Table 4
Distribution of Controlled Sample Groups on Duration of Illness

Months duration	Improved group	Unimproved group
Under 6	12	12
6 – 11	12	17
12 – 23	18	11
24 – 35	13	20
Over 36	35	30
Chi square	4.40	

long duration of illness, thus raising the mean for the group.

The net result would seem to be that the elimination of the effect of the other variables significantly related to outcome also eliminated the relationship obtaining between duration and outcome. And it would appear that the importance of duration as a prognostic sign is derived chiefly from its relationship to other variables of more basic importance that are significantly related to outcome. It would appear that the safest conclusion that could be reached is that the mere passage of time is in itself neither good nor bad so far as the possibility of "curing" a mental illness is concerned, but rather it is the processes that are allowed to continue unchecked during this passage of time that are of importance to the outcome of a mental illness.

Summary

It was the purpose of this study to determine the relationship between duration of mental illness and the outcome of illness when the effect of other variables, significantly related to outcome, had been controlled.

A sample of 486 patients was used. The relationship between outcome of illness and 21 variables was determined. Then the group was divided into an improved and an unimproved group, and cases dropped from each of the groups until there were no significant differences between the groups on any of the variables other than duration of illness.

It was found that after this controlling was completed there were no significant differences between the groups on duration of illness. This result was interpreted as indicating that the prognostic importance of duration of illness derived from its relationship to other variables related to the basic disease process, that the mere passage of time presaged neither good nor ill for mental illness, but rather, that it was the processes allowed to progress unchecked during this period that were of prognostic importance in mental illness.

Received February 23, 1954.

References

1. Kalinowsky, L. B., & Hoch, P. *Shock treatment and other somatic procedures in psychiatry.* New York: Grune & Stratton, 1946.
2. Pascal, G. R., Swensen, C. H., Feldman, Dorothy A., Cole, Mary E., & Bayard, Jean. Prognostic criteria in the case histories of hospitalized mental patients. *J. consult. Psychol.*, 1953, **17**, 163-171.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Rigidity in Conceptual Thinking in Exogenous and Endogenous Mentally Retarded Children

J. G. McMurray
University of Toronto[1]

It is noted occasionally by clinicians that exogenous mentally retarded children, as compared with endogenous, are significantly less able to perform such memory-for-designs tests as that found at year IX on the Revised Stanford-Binet Intelligence Scale. At least one investigator [4] has suggested that their inferior performance is not due to faulty memory, as is frequently supposed, but to other factors. In an earlier paper [15] the author attempted to demonstrate that the unusual visual perception of brain-injured subjects is one factor accounting for their relative inability to reproduce visual designs. The observation holds true even when the stimulus figures are to be reproduced from copy and not from memory. The exogenous subjects' distorted reproductions of visual designs appear not to be the result of faulty memory but rather of misperception of the stimulus figures.

Using the Ellis Visual Designs Test [21] as a means of quantifying a child's drawing ability, the author noted earlier [14] that some mentally retarded children exhibited difficulty in changing their responses appropriately when shifting from one drawing to the next. There was a tendency for some of the subjects to produce the same or similar drawings in response to radically different stimulus figures, which might be called a tendency to perseverate. These perseverative responses seemed to be more common for the brain-injured than for the familial defectives. It was therefore thought possible that a further factor contributing to the defective drawing performance of the exogenous subjects might be their greater tendency toward rigidity in conceptual thinking.

In accordance with Goldstein [5], Werner [20], and others, rigidity in the present study is defined as the tendency of an individual to resist change maladaptively. Perseveration is regarded as one type of rigid response that repeats itself maladaptively. Flexibility is regarded as the capacity of an individual to alter mental sets and corresponding responses appropriately as these are required by changes in the stimulus situation.

A number of investigators have found evidence that the brain-injured are characterized by rigidity. In 1928 Cameron [2] showed that brain-operated rats are less able than healthy animals to break a habit and form a new one. Maier [16] has demonstrated that operated rats tend to perseverate, that is, to repeat the same errors in successive trials. Goldstein [6, 7], in more recent studies with brain-injured and normal adults, concluded that rigidity is a general pathological trait that is present irrespective of the focus of lesion.

Rigidity is not, however, an exclusive characteristic of the brain-injured. It has been demonstrated that mentally retarded individuals whose deficiency is endogenous also show rigidity. Werner [19] refers to their rigidity as *subnormal,* in comparison with that of the brain-injured child which is termed *abnormal* rigidity. One example of a study made of subnormal rigidity is that of Lewin [12], who found that endogenous as compared with normal subjects displayed a significantly higher degree of rigidity. The investigations of Strauss and Werner [18] and Carrison and Werner [3], dealing with abnormal rigidity, have shown that exogenous as compared with endogenous defective children exhibit perseveration and a general difficulty in shifting from one task to another.

Many research workers, including Goldstein and Scheerer [8], Weigl [19], Kasanin and Hanfmann [11], and Werner [19, 20], have devised tests in an attempt to measure the presence or absence of rigidity. After exploring the applicability of these tests on small groups of mentally defective children, the author found them generally inappropriate. The instrument finally selected for use in the present study was a modification of a Weigl-type card-sorting test first described by Berg [1], and later by Grant and Berg [9], at the University of Wisconsin. The test is an objective technique for measuring human abstraction and shift of set and is claimed by its authors to yield quantitative measures of these aspects of thinking and reasoning.

---

[1] The writer gratefully acknowledges the cooperation of the superintendent and staff of the Ontario Hospital School, Orillia, Ontario, for making available the facilities for this study.

The Investigation

Problem. The purpose of this study is to investigate the hypothesis that exogenous as compared with endogenous mentally retarded children show greater rigidity in conceptual thinking.

Subjects. Two groups of mentally defective subjects, exogenous and endogenous, more fully described in a previous publication [15], were chosen from a total population of 2,000 institutionalized patients. The 15 exogenous subjects included all of the available patients who satisfied the rigid selection criteria. The 15 endogenous subjects, six boys and nine girls, were paired as closely as possible with the exogenous subjects for sex, chronological age, and intelligence. Intelligence was measured using the Revised Stanford-Binet Intelligence Scale, Form M. The two groups were composed of subjects between the ages of 10 and 20 with IQ's between 40 and 75. Subjects were selected by a member of the medical staff, a geneticist, and the author. The presence or absence of brain injury in each subject was established (*a*) by examination of case history material and (*b*) by neurological examinations of the subjects themselves.

Method. The modified Wisconsin card-sorting test used in the present study consists of a pack of 27 response cards made of white cardboard and approximately 3½ × 5 inches in size. Each card contains from one to three identical figures of a single color. Three kinds of figures—diamonds, squares, and triangles—and three colors—green, blue, and red—are used. A single card might have one green diamond, or two blue squares, or three red triangles, or any such combination. Thus, the cards may be sorted according to any one of three different principles (*a*) the color, (*b*) the form, or (*c*) the number of figures on each card. In addition to the 27 response cards which are sorted by each subject, there are three stimulus cards which are placed on the table by the examiner as guides for sorting. The three stimulus cards are always placed in the following order: the card showing one green diamond; the card showing two blue squares; and the card showing three red triangles.

When the cards are first presented to the subject for sorting, they are arranged in a standard order, but for any subsequent sortings the cards are simply shuffled. Sorting in terms of color is the principle first accepted as correct. The subject, of course, receives no instructions to this effect. He is simply told that his response is incorrect when he is not placing the response cards in front of the stimulus cards according to color. After a subject has sorted five successive cards according to the principle of color it is assumed that he has grasped the principle that the examiner accepts as correct and is using it in his sorting. At this point the examiner calls such sortings of the cards incorrect and continues to do so until the subject next learns to sort them according to either the form or the shape of the designs on the cards. The same general procedure is repeated until the three sorting principles of color, form, and number have been successfully grasped or until the pack of response cards has been sorted completely six times. The following instructions are given to each subject:

"I want you to put these cards on the table in front of one of the three cards already there (examiner demonstrates). Lay the cards down one at a time. As you can see, the cards differ in three ways: in the *color* of the spots on each card; in the *number* of spots on each card; and in the *shape* of the spots on each card. Put each card down where you think it should go and I will tell you when you are right and when you are wrong. Remember that there are three ways to sort the cards. Don't take too long but get as many right as you can."

Scoring procedure. Rigid behavior on the card-sorting test is scored according to three different criteria:

1. Perseverative responses: A perseverative response occurs when a card is sorted in terms of a principal previously accepted as correct (e.g., a card sorted according to the principle of color when it is supposed to be sorted according to the principle of form).

2. The total number of cards sorted (correct-

ly or incorrectly) before each of the three sorting principles is grasped.

3. The time required to grasp and apply each new sorting principle.

Results and Discussion

The exogenous as compared with the endogenous mentally retarded subjects showed, on the average, significantly more rigid behavior on the modified Wisconsin card-sorting test. As indicated in Table 1, rigidity was apparent not only in the subjects' total number of perseverative responses, but also in the other two measures of sorting efficiency, that is, in the total number of cards sorted and in the time needed to grasp and apply the three sorting principles.

Table 1

Group Difference in Total Rigidity Scores on the Modified Wisconsin Card-Sorting Test

Marking criteria	Endogenous subjects ($N=15$)	Exogenous subjects ($N=15$)	r†
Color			
Perseverative score	0	0	
Number of cards sorted	198	311	1.56
Time in seconds	1544	2917	1.80
Form			
Perseverative score	23	160	2.55*
Number of cards sorted	210	580	3.70**
Time in seconds	1551	3971	2.95*
Number			
Perseverative score	52	248	4.91**
Number of cards sorted	372	915	4.71**
Time in seconds	2719	5993	2.91*
Total scores			
Perseverative score	75	408	5.38**
Number of cards sorted	780	1793	6.17**
Time in seconds	5715	13035	4.31**

† The t ratios were calculated using Weaver's variation [13] of the usual formula for small matched groups.
* Significant at 5% level of confidence.
** Significant at the 1% level.

As mentioned in the section on scoring procedure, the perseverative scores represent the number of errors made when cards are sorted in terms of the principle previously accepted as correct. This interpretation of perseveration agrees with that of Werner [19], who uses the term to indicate the persistent repetition or continuation of an activity once begun. A direct comparison of the perseveration scores of the two etiological groups showed the endogenous group mean to be 5.0 (range 2—12) and the exogenous group mean to be 27.2 (range 2 — 59). In addition to the differences between the scores of the two groups being statistically significant at the one per cent level of confidence, there was an apparent cut-off score. All of the endogenous, but only three of the exogenous, subjects had a score of 12 or less. The group differences are presented graphically in Figure 1. As was expected, a definite etiological difference in favor of the endogenous mentally retarded subjects appears to obtain.

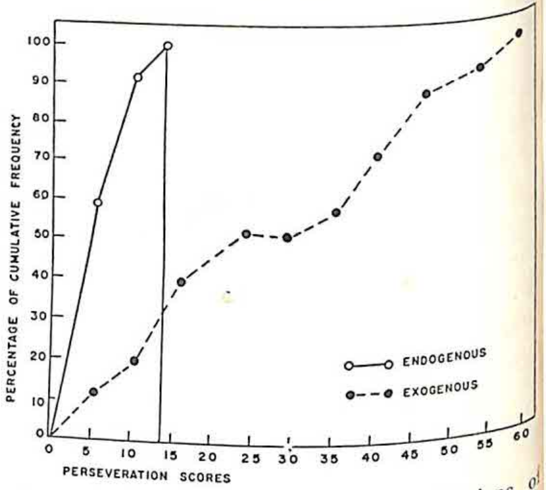


Fig. 1. Cumulative percentage distributions of perseveration scores for the endogenous and exogenous groups.

In addition to the greater number of perseverative responses made by the exogenous subjects, both the number of responses necessary and the time taken to achieve satisfactorily the three principles were significantly greater for the exogenous than for the endogenous mentally defective subjects. The longer time and greater number of responses required is considered by the author to be the result of the exogenous subjects' reduced ability to shift their "set" readily from one sorting principle to another.

In comparing the scores achieved by the two groups of subjects on the card-sorting test, it was found that both groups encountered greater difficulty as the test progressed. While it has already been shown in Table 1 that the exogenous, as compared with the endogenous, mentally retarded subjects showed more rigid behavior in the test as a whole, the groups responded similarly in that they both found sorting according to the principle of number more difficult than sorting according to the principle of form. This is probably due, in part at least, to the fact that rigidity is proportional to the number of principles by which the subjects have learned to sort the cards. The results of the present study, however, tend to support Heidbreder's [10] study performed with normal subjects, which found that of the three concepts, number was the last concept to determine a sorting. The present study is also in agreement with the investigations reported by Senden [17] on the perceptions of the congenitally blind after operation. He indicates that the concept of color persistently dominates form in the first vision of such patients. Eleven months after operation the color names that had been learned were retained, but the little that had been learned of form was forgotten.

A comparison of the rigidity scores in this investigation with the perception scores obtained in a previous study [15] denoted no significant relationship between these factors. Product-moment correlations (endogenous group + .20; exogenous group — .10) were not significant for either etiological group. This would tend to indicate that visual perception, as measured by Lissajou's figures, is relatively independent of a subject's ability to shift his set from one task to another on a card-sorting task.

Summary and Conclusions

Previous investigations [5, 14] have indicated that exogenous mentally retarded children are significantly less able than those of endogenous etiology to reproduce visual designs from copy. In a previous paper [15] the author provided evidence that the unusual visual perception of the exogenous defectives was one factor accounting for their poor drawing performance. The purpose of the present study was to investigate the hypothesis that the relative inability of the brain-injured to reproduce visual designs from copy, in comparison to familial defectives, is due to rigidity in conceptual thinking.

In the present study a modified version of the Wisconsin card-sorting test was employed to study rigidity in conceptual thinking in exogenous and endogenous mentally retarded children. The subjects from the two groups were paired on the basis of sex, chronological age, and intelligence.

It was demonstrated that the brain-injured are significantly more rigid in conceptual thinking as shown by their greater tendency to perseverate on a card-sorting task. The exogenous child is indeed less able to shift his mental set from a sorting principle such as color to another such as form or number. Such findings warrant the conclusion that the relatively poor drawing performance of exogenous defectives on such memory-for-designs tests as that found at year IX on the Binet are not due to faulty memory but to such factors as perseveration and rigidity in conceptual thinking.

Received January 14, 1954.

References

1. Berg, Esta A. A simple objective technique for measuring flexibility in thinking. *J. gen. Psychol.*, 1948, **39**, 15-22.
2. Cameron, N. Cerebral destruction in its relation to maze learning. *Psychol. Rev.*, 1928, **39**, 68-71.
3. Carrison, D., & Werner, H. Principles and methods of teaching arithmetic to mentally retarded children. *Amer. J. ment. Def.*, 1943, **47**, 281-289.
4. Cassel, R. H. Relation of design reproduction to the etiology of mental deficiency. *J. consult. Psychol.*, 1949, **13**, 421-428.
5. Gelb, A., & Goldstein, K. *Psychologische Analysen hirnpathologischer Faelle.* Leipzig: Barth, 1920.
6. Goldstein, K. The modifications of behavior consequent to cerebral lesions. *Psychiat. Quart.*, 1936, **10**, 586-610.
7. Goldstein, K. *The organism.* New York: American Book Co., 1939.
8. Goldstein, K., & Scheerer, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
9. Grant, D. A., & Berg, Esta A. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *J. exp. Psychol.*, 1948, **38**, 404-411.
10. Heidbreder, Edna. Conceptual achievements during card sorting. *J. Psychol.*, 1949, **27**, 3-39.
11. Kasanin, J., & Hanfmann, Eugenia. An experimental study of concept formation in schizophrenia. *Amer. J. Psychiat.*, 1938, **95**, 35-48.

12. Lewin, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.

13. McCullough, M. W. Wechsler-Bellevue changes following prefrontal lobotomy. *J. clin. Psychol.,* 1950, 6, 271-274.

14. McMurray, J. G. A study of factors contributing to poor drawing performance in exogenous as compared with endogenous feebleminded children. Unpublished master's dissertation, Univer. of Western Ontario, 1952.

15. McMurray, J. G. Visual perception in exogenous and endogenous mentally retarded children. *Amer. J. ment. Def.,* 1954, 58, 659-663.

16. Maier, N. R. F. The effect of cerebral destruction on reasoning and learning. *J. comp. Neurol.,* 1932, 54, 45-75.

17. Senden, M. *Raum- und Gestaltauffassung be operierten Blindgeborenen vor und nach de Operation.* Leipzig: Barth, 1932.

18. Strauss, A. A., & Werner, H. Experimenta analysis of the clinical symptom "perseveration" in mentally retarded children. *Amer. J. ment. Def.,* 1942, 47, 185-188.

19. Werner, H. Abnormal and subnormal rigidity *J. abnorm. soc. Psychol.,* 1946, 41, 15-24.

20. Werner, H. The concept of rigidity: a critical evaluation. *Psychol. Rev.,* 1946, 53, 43-52.

21. Wood, Louise, & Schulman, Edythe. The Ellis visual designs test. *J. educ. Psychol.,* 1940, 31, 591-602.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Test Performance of Preschool Children With and Without Organic Brain Pathology[1]

Miriam Stearns Haines
The Public Schools, Montclair, New Jersey

The diagnosing of brain injury is of major importance. Until recently the responsibility for this diagnosis rested solely with the medical profession. Today it is recognized that psychological findings may contribute to the final diagnosis along with the neurological examination and such laboratory tests as the electroencephalogram. Although a correct diagnosis is important for mature individuals, it is of even greater consequence for young children because of the impact of brain injury on the developing organism [4]. Brain-injured adults retain much of what they have learned even though the acquiring of new skills is impaired. The situation is more serious, however, for children who have sustained cerebral insult before many tasks are normally learned. If a correct diagnosis is not made, they may not receive the specialized training which they need in order to learn normally expected skills. Methods have been especially designed to teach them such necessary functions as reading and arithmetic fundamentals [16]. If a diagnosis of primary amentia is made rather than of secondary amentia due to brain injury, these children are frequently placed in institutions for the endogenous mental defective and barred from an opportunity for even rudimentary academic learning [5].

Two assumptions are made when a psychological test is expected to differentiate between young children with and without brain injury: (*a*) that children's abilities are sufficiently developed and specialized at an early age for differential patterns to emerge, and (*b*) that the test is sensitive enough to tap varying degrees of ability [2]. The fact that test items in both verbal and performance areas are administered to children and infants as young as four weeks on the Gesell Developmental Schedules and at two months on the Cattell Infant Tests lends weight to the concept of differential abilities in the young child [1, 9].

On the basis of previous research with older children there is reason to suspect that young brain-injured children have difficulty in conceptual thinking, in visual-motor and perceptual judgment, and there is also the possibility of more general mental retardation [7, 13, 14, 17, 21]. Whether or not intellectual impairment follows a brain injury depends on the area of the brain affected. If the lesion is in the cerebral cortex, intellectual impairment or "deficit" is a definite possibility although even here not an inevitable consequence [18]. Literally, any case of feeblemindedness might be considered as possibly due to birth lesion when one considers the whole process of fetal development and the difficulties of childbirth.

Method

The instrument chosen in this investigation to test varying degrees of ability was the Merrill-Palmer Scale of Mental Tests [19]. It was selected primarily because it consists of a battery of verbal and performance tests that permits scoring for different degrees of ability through the preschool years. In addition, it has strong appeal for young children and very few resist "playing the games." Motivation is an important factor since cooperation is frequently difficult to obtain in brain-injured children on tests without intrinsic appeal. In fact, electroencephalograms are often reported to be unsatisfactory in these children because of lack of cooperation.

It is hypothesized that young children with brain injury will show decreased efficiency on tests involving conceptual thinking, visual-motor skill, and perceptual judgment. Specifically, it is expected that the brain-injured group should show significantly lower means on the selected Merrill-Palmer tests that purport to measure these abilities than the control group. A brain-injured child in this study is defined as one who has suffered a brain lesion due to infection or injury occurring before, during, or after birth. Conceptual thinking is conceived as the ability to form a concept, to classify according to a common factor. Visual-motor skill is dependent upon eye-hand coordination, finger dexterity, speed of reaction time, spatial evaluation, and visual memory. Perceptual judgment implies the discriminative observation of differences, relationships, and organizations.

---

[1] Acknowledgment is made to the National Institutes of Health of the U. S. Public Health Service for a fellowship grant to the author to conduct this study and to Miss Louise R. Hewson, Director, Psychology Department, Neurological Institute, New York City to collect the data.

The experiment was designed for a brain-injured group of 100 children and two non-brain-injured groups consisting of 100 behavior-problem and 100 foster-home children to whom the Merrill-Palmer and the Revised Stanford-Binet, Form L, were administered. The selection of behavior-problem children for comparison with the brain-injured group was made partly because of their availability in the clinic as a non-brain-injured population but mainly because differentiation between children with brain injury and those with behavior problems is an especially important issue which the neuropsychiatrist refers to the psychologist [3, 10, 11, 12]. Comparison could have been made with so-called "normal" children in nursery school, kindergarten, and primary grades, but the foster-home group was selected instead because from the standpoint of this study there would have been several disadvantages with a normal group. First, there is screening of children for admission to these educational groups, and those with low intelligence quotients would probably have been eliminated, especially where there were physical stigmata. Second, the testing could not have been done by the same examiners at the same place, whereas identical conditions prevailed for the three groups that were selected.

The only criteria for selection of subjects besides their diagnosis were chronological age from 3 years, 0 months to 7 years, 10 months; Intelligence Quotients of at least 45 on the Revised Stanford-Binet, Form L; and basal age at least 30 months on the Merrill-Palme [20]. These limitations stem from the selectio of Merrill-Palmer tests and a choice of a predominantly preschool population. The mea Stanford-Binet IQ for the brain-injured an foster-home children was found to be 93 an for the behavior-problem group 89. Althoug no attempt had been made to match the thre groups for sex, the brain-injured were found t have a one-to-one ratio, while both the non-brain-injured groups had a higher proportio of boys. None of the children had visual o auditory impairments, nor was there a difference in the socioeconomic backgrounds.

The etiologies in the exogenous group were: birt injury, 5; meningitis or encephalitis, 15; hypoplasi of the brain, 23; tumor of the brain, 10; grand o petit mal epilepsy, 16; cortical seizures, 11; hydrocephalus, 9; brain-injury etiology unknown, 11. Th electroencephalogra and pneumoencephalogra were abnormal. The behavior-problem and foste home children are known as non-brain-injure groups because their developmental histories an neurological examinations gave no evidence of lesion of the central nervous system.

The tests selected were the Action Agent, th Buttons, the Copying a Circle, Cross, and Star, th Seguin Formboard, the Pyramids, the Pink Towe the Mare and Foal, and the Picture Puzzles becaus they tap a variety of abilities as well as provide fo varying degrees of proficiency among children fro 30 to 77 months. The tests usually administered below 30 months were omitted because they are base on the all-or-none principle of scoring; for exampl the child can either cut with a pair of scissors or h cannot; he can either identify himself in a mirro or he cannot.

An analysis of the abilities in the Merrill-Palmer was made as a preliminary step so tha the abilities involved in each test were define prior to the analysis of the relevant data. Thre competent psychologists who had clinical experience with the Merrill-Palmer scale and ha taught and demonstrated its technique to graduate students and externs were chosen to chec on the writer's analysis of abilities. The judge were sent the writer's analysis of the abilitie involved in the Merrill-Palmer tests togethe with a list of definitions for the abilities. Th combined efforts resulted in a grouping of th tests in terms of the abilities in each test as they related to the hypothesis under investigation. For example, under conceptual thinkin was listed the test which involves this abilit primarily; under visual-motor skill those tests which embody spatial evaluation, speed of reaction time, and visual memory; and under perceptual judgment tests which are predominantly dependent upon this ability.

Results

The Merrill-Palmer mean scores for each of the three diagnostic groups ranged from 3.0 to 4.8, where there is a possible range from 0 to 7.0. When the *t* test of the significance of the differences between the means of the three groups, for each subtest, was computed, it was found that no test was significantly different at the .05 level in all instances. When the *t* test was applied to all differences between subtest means within each group, no one test was found to reach significance at the .05 level in all seven instances. Thus, the null hypothesis must be accepted in terms of the criterion, and the differences among the three groups and between tests in any one of the groups must, therefore, be attributed to chance. Table 1 shows the *t* tests between the means of the eight Merrill-Palmer tests for each pair of diagnostic groups.

Discussion

The results of this investigation do not demonstrate that young children with brain injury show decreased efficiency on tests presumably measuring conceptual thinking, visual-motor skill, and perceptual judgment. This may be because abilities at this young age level are general rather than specific [8]. Other studies which report significant differences dealt with older children and adults. It is also possible that the Merrill-Palmer as a measuring device is not a sensitive enough instrument to differentiate degrees of conceptual thinking, spatial evaluation, and perceptual judgment compared to those tests which have been specifically designed for differential diagnosis. The fact that little relationship was found between Stanford-Binet IQ's and Merrill-Palmer test scores tends to support this interpretation. The present research is in line with the recent attempts to explore the contributions of intelligence tests to neuropsychiatric diagnosis. Since these tests were designed to measure global intelligence and are based on theory directed toward that end, it should be no disparagement of the Merrill-Palmer that it does not perform a function foreign to its basic premise [15].

Summary

This study investigated the ability of selected subtests of the Merrill-Palmer scale to differentiate between brain-injured and non-brain-injured children, three to seven years of age. The experimental group of 100 brain-injured children was compared to two control groups of 100 behavior-problem children and 100 foster-home children. Analyses were made by *t* tests of the differences of mean subtest scores within each group. Although a few scattered *t* tests showed statistical significance, no test pattern emerged that validly differentiated the brain-injured from the other groups. The null hypothesis must be accepted. It is concluded that subtests of the Merrill-Palmer scale are not useful for the differential diagnosis of brain-injured children from behavior-

Table 1

t Tests between Means of Eight Merrill-Palmer Tests for Each Pair of Diagnostic Groups

Groups	Tests							
	Seguin	Picture Puzzles	Buttons	Action Agent	Pink Tower	Mare & Foal	Pyramids	Circle, Cross, Star
Brain-injured vs. behavior problem	1.27	1.66	.52	1.76	1.84	.90	1.74	.92
Behavior problem vs. foster home	2.21*	2.70**	.23	3.02**	1.49	.83	1.49	1.60
Brain-injured vs. foster home	.88	1.01	.78	5.06**	.30	.06	.22	.58

* Significant at .05 level of confidence (1.96).
** Significant at .01 level of confidence (2.576).

problem or foster-home children.

Received January 28, 1954.

References

1. Cattell, Psyche. *The measurement of intelligence of infants and young children.* New York: Psychological Corp., 1947.
2. Clark, M. P. Changes in primary mental abilities with age. *Arch. Psychol.*, 1944, No. 291.
3. Cobb, S. Personality as affected by lesions of the brain. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944. Pp. 550-581.
4. Doll, E. A. Behavioral consequences of cerebral birth lesions. *Amer. J. Obst. Gynecol.*, 1936, **31**, 866.
5. Fay, Temple, & Doll, E. A. Organic impairment simulating mental deficiency. *Amer. J. Orthopsychiat.*, 1949, **19**, 112-119.
6. Fisher, R. A. *Statistical methods for research workers.* (6th Ed.) London: Oliver & Boyd, 1936.
7. Frazeur, Helen A., & Hoakley, Z. Pauline. Significance of psychological test results of exogenous and endogenous children. *Amer J. ment. Def.*, 1947, **51**, 384-388.
8. Garrett, H. A developmental theory of intelligence. *Amer. Psychologist,* 1946, **1**, 372-378.
9. Gesell, A., & Amatruda, Catherine S. *Developmental diagnosis.* (2nd. Ed.) New York: Hoeber, 1948.
10. Halstead, W. C. *Brain and intelligence.* Chicago: Univer. of Chicago Press, 1947.
11. Kasanin, J. Personality changes in children following cerebral trauma. *J. nerv. ment. Dis.*, 1929, **69**, 385-406.
12. Lashley, K. S. *Brain mechanisms and intelligence.* Chicago: Univer. of Chicago Press, 1929.
13. Sarason, S. B. *Psychological problems in menta deficiency.* New York: Harper, 1949.
14. Sarason, S. B., & Sarason, Esther K. The discriminatory value of a test pattern with cerebral palsied defective children. *J. clin. Psychol.*, 1947, **3**, 141-147.
15. Shaffer, L. F. Theoretical bases for psychometric tests. In P. H. Hoch & J. Zubin (Eds.), *Relation of psychological tests to psychiatry.* New York: Grune & Stratton, 1950. Pp. 26-38.
16. Strauss, A. A., & Lehtinen, Laura E. *Psychopathology and education of the brain-injured child.* New York: Grune & Stratton, 1947.
17. Strauss, A. A., & Werner, H. Disorders of conceptual thinking in the brain-injured child. *J. nerv. ment. Dis.*, 1942, **96**, 153-172.
18. Strecker, E. A., & Ebaugh, F. G. Neuropsychiatric sequelae of cerebral trauma in children. *Arch. Neurol. Psychiat.*, 1924, **12**, 443-453.
19. Stutsman, Rachel. *Mental measurement of preschool children.* Yonkers, N. Y.: World Book Co., 1931.
20. Terman, L. M., & Merrill, Maud A. *Measuring intelligence.* Boston: Houghton Mifflin, 1937.
21. Werner, H., & Strauss, A. A. Impairment in thought processes of brain-injured children. *Amer. J. ment. Def.*, 1943, **47**, 291-295.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Sex Differences on the Revised Beta Examination

George Lavos
Michigan School for the Deaf

The Revised Beta Examination is a nonlanguage test measuring "general intellectual ability" for those 16 years of age and older [1, p. 2]. No reports of its use with the deaf appear in the literature although it is adaptable for use with the deaf in administration, format, and nonlanguage response.

The Problem

The Revised Beta Examination is provided with norms based on white male adults. These men were selected so that they would represent proportionately in the normative population the educational and socioeconomic status of the adult white males in the 1940 U.S. census. Although the standardization group is limited to men, the Manual states that "the Examination has been used for testing women as well as men for many years and no report has been found indicating that different norms were required for the different sexes" [1, p. 6].

The assumption that the norms based on men are appropriate for women should be tested before use of the examination is recommended for deaf women and older girls. The problem is, then, when used with deaf youths is there a sex difference in the intelligence quotients based on the Revised Beta Examination that cannot be attributed to chance sampling in the population?

Method

The records of 188 deaf pupils, between the ages of 16 and 19 at the time of testing, were available. The average age at the time of administering the test was 17-3. The pupils were accustomed to testing since the Beta test was the third group test of intelligence used in the school from which the data were obtained. There were 79 girls and 109 boys in the group. All were regularly enrolled pupils in the school and had severe hearing losses.

The intelligence quotients for the total group ranged from 62 to 128. These quotients were calculated according to the Lindner-Gurvitz procedure recommended for use with the Beta test.

In order to find out whether there is a sex difference on the Beta, it is necessary to ascertain whether the performance of the two sexes is the same on similar tests. Since the Beta was the third of a series of tests of nonlanguage intelligence used with the subjects of the study, results of the first two tests could be used to assess the possibility that a sex difference already exists in the group. The critical ratio of the difference between the means of the quotients for boys and girls on the first of the two tests given earlier, the Chicago Non-Verbal Examination, was 1.44. The critical ratio for the mean quotients from the second test, the Pintner Non-Language Test of Mental Ability, Intermediate Series, was .005. Both ratios indicated that no sex differences exist on the Chicago and the Pintner tests that cannot be attributed to chance factors.

Results

The mean Beta intelligence quotient of the girls (94.4) was lower than that of the boys (103.0). The critical ratio of the difference in means was 4.34. The probability is less than 1 in 1,000 that a critical ratio this large can arise by chance. There is a reliable difference in the means of deaf boys and deaf girls on the test; consequently, the two sexes cannot be considered to be samples from a common population with respect to the capacities measured by the test.

In the light of this finding it is desirable to see on which subtests of the Beta deaf boys and girls differ significantly in performance. By using the weighted scores for each subtest that are provided in the Lindner-Gurvitz standardization, it is possible to compare directly any person within the age range from 16 to 19, 20 to 24, and so forth, without considering growth. Thus the boys and the girls, each group ranging in age from 16 to 19, can be directly compared without regard to variations in age within that range.

The differences in means of the weighted scores for the boys and the girls on each of the six subtests of the Beta were tested for significance. Three mean differences were statistically significant, and these were in favor of the boys. The three subtests in which significant differences appeared were "error recognition," "paper form board," and "picture completion."

The subject matter of the drawings in the items for "error recognition" is largely masculine. The "paper form board" subtest is similar to those appearing in mechanical aptitude tests. In "picture completion" there is no preponderance of masculine or feminine orientation.

Summary

In order to test the assumption that the norms for the Revised Beta Examination can be used with deaf women, the norms being derived from hearing males only, the test result of 79 deaf girls and 109 deaf boys between 1 and 19 years of age were analyzed. It wa shown that the performance of these subject was not indicative of a sex difference in intelligence quotient on the Chicago Non-Verbal Examination and the Pintner Non-Languag General Ability Test, which had previousl been given to the subjects. On the Beta, however, a significant sex difference in mean inteligence quotient existed in favor of the boys Subtest analysis located the difference in three subtests, two of which are masculine in orientation. It is concluded that the Beta test cannot be used with deaf girls of 16 to 19 years o age without appropriate norms.

Received February 9, 1954.

Reference

1. Lindner, R. M., & Gurvitz, M. *Revised Be Examination manual.* New York: Psychologic Corp., 1946.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

A Validation of a Bender-Gestalt Scoring System[1]

Murray Lonstein
VA Hospital, Lebanon, Pennsylvania

The Bender-Gestalt test (B-G) has become a part of test batteries in many hospitals and clinics throughout the country. Its need for quantification and validation has been recognized, and, accordingly, scoring systems have been devised by several investigators [1, 4, 5]. One such scoring system, constructed by Pascal and Suttell [4], appears worthy of further study in view of the fact that its authors have claimed that the method discriminates normal subjects from neurotics and psychotics. Furthermore, Robinson [6] found that this system differentiates the B-G performances of paretics and schizophrenics, and Cornutt [2] supported the claims of Pascal and Suttell in regard to scoring and collection of their normative data.

The present study was designed to validate further the Pascal-Suttell scoring system and to establish its practical usefulness in a hospital setting. It was held that even if the system is not sensitive enough to distinguish neurotic from normal subjects, as Gobetz [3] claimed, its efficacy in separating psychotic from nonpsychotic patients would merit its continued use.

Procedure

A total of 96 B-G records of psychiatric patients in a veterans hospital was scored by the author according to the directions given by Pascal and Suttell. Briefly, this technique involves an analysis by the psychologist of Designs 1-8, which are scored according to criteria given in their manual. The number of points for each specified deviation increases with its severity. The sum of these points is the raw score, which is converted into a standard measure or *z* score.

After the records were scored, the psychiatric diagnoses of the patients were obtained and the records were placed into two groups, "psychotic" and "nonpsychotic." In the psychotic group there were 41 schizophrenics, 4 psychotics with organic brain disease, and 2 manic-depressives. In the nonpsychotic group there were 40 anxiety reactions, 6 passive-aggressive reactions, 1 obsessive-compulsive reaction, 1 dissociative reaction, and 1 alcoholic.

Since the Pascal-Suttell *z* scores are based on norms that exclude subjects whose education is below the ninth grade, 23 patients who did not meet this standard were not used in the *z*-score comparison. They were included, however, in the raw score analysis.

Results

As indicated in Table 1, the means of the standard score distributions of the two groups differed significantly; likewise, the means of the larger raw score distributions also differed significantly. These findings indicate that the Pascal-Suttell scoring system meets the statistical criterion in differentiating psychotic from nonpsychotic patients, using either raw or standard scores.

Table 1

Statistical Comparisons of B-G Standard and Raw Scores of Psychotic and Nonpsychotic Groups

Statistic	Standard scores		Raw scores	
	Psychotic	Non-psychotic	Psychotic	Non-psychotic
N	36	37	47	49
M	72.19	56.05	35.30	23.96
SD	22.48	14.12	22.56	14.53
CR	3.67 ($p < .001$)		2.92 ($p < .01$)	

Although the null hypothesis may be rejected with a high degree of confidence, it was deemed more appropriate to evaluate this technique as a practical method by determining from the data the percentage of true and false positives. In order to satisfy this important criterion, it was necessary to find appropriate cutoff points to separate the two groups. Pascal and Suttell report that *z* scores between 50 and

---

[1] From the VA Hospital, Lebanon, Pennsylvania.

72 are suspect for psychiatric illness. A *z* score of 60, one standard deviation above their normal group mean of 50, is suggested by them as a reasonable cutoff score in differentiating between patients and nonpatients.

Inspection of the data reveals that of the 36 psychotic records on which *z* scores were computed, 24 patients had scores greater than the suggested cutoff score of 60, and 12 had scores below 60. As indicated in Table 2, 66.66 per cent of the psychotics were correctly identified. Of the 37 nonpsychotic patients, 13 had scores above 60, whereas 24 had scores below 60; that is, 64.86 per cent were correctly identified. In the total sample of 73 patients, the percentage of true positives was 65.76 and the percentage of false positives was 34.24.

Table 2
Percentage of True and False Positives with Cutoff B-G Scores of 60 and 65

Percentages	Cutoff score at 60	Cutoff score at 65
% psychotics correctly identified	66.66	63.89
% nonpsychotics correctly identified	64.86	73.38
Total % false positives	34.24	28.77
Total % true positives	65.76	71.23

Since all the subjects in this study were hospitalized and therefore maladjusted to some degree, it seemed reasonable to select a somewhat higher cutoff score; accordingly, a cutoff score of 65 emerged as a more effective score for reducing the total percentage of false positives. Only 8 of the 37 nonpsychotic records had *z* scores over 65, and 13 of the 36 psychotics had scores below 65; that is, with the higher cutoff score, 1 additional psychotic was misidentified, but 5 more nonpsychotics were correctly identified. The percentages in Table 2 reflect this trend and show the total percentage of true positives as 71.23, and the total percentage of false positives as 28.77. In other words, the technique correctly separated psychotics from nonpsychotics in about 7 out of 10 cases when a cutoff score of 65 was used.

Since all of the B-G performances were timed when the test was administered, it was believed that knowledge of possible differences between the two groups on the time variable would be helpful diagnostically. Accordingly the time measures (seconds taken to copy al of the designs) were compared on a group basis The psychotic group's mean time measure wa 445.03 seconds, with a sigma of 290.25 se onds; the nonpsychotic group averaged 331.0 seconds, with a sigma of 112.82 seconds. Th critical ratio derived in this comparison wa 2.29, significant beyond the .05 level of con fidence—that is, psychotic patients worke more slowly on the test than did nonpsycho tic patients.

Discussion

Although the Pascal-Suttell scoring system meets a high statistical level of confidence i differentiating between hospitalized psychoti and nonpsychotic patients, it is by no mean infallible in the practical, clinical setting Nevertheless, in practical work it can discri minate between these two groups on a bette than chance basis. The variables selected fo scoring are therefore of diagnostic importan and perhaps should be studied by those wh use this test even though they may not formal score each B-G record. However, since th scoring procedure can be learned easily and record can be scored quickly, the techniqu should be of practical usefulness in clinic settings.

As Pascal and Suttell point out, an indiv dual *z* score derived by use of their system mu be interpreted with caution. Although a high *z* score suggests psychiatric illness, it does no always mean that the subject is psychotic. Som factors incidental to the testing situation, such as fatigue, may contribute to high scores. Lik all such test measures, the *z* score must b viewed along with other test scores and clinica data. Furthermore, the fact that the criterio used in this study is psychiatric diagnosis im poses some limitations on the reliability of thes results.

Summary

A validating study was carried out on the Pascal-Suttell scoring system of the Bender Gestalt test. The records of 96 psychiatric patients were scored, and it was found that this technique differentiated psychotic from non psychotic records by meeting statistical criteria and, in a limited way, practical criteria. An in-

cidental finding is that psychotic patients took more time in reproducing the designs than did nonpsychotic patients.

Received December 16, 1953.

References

1. Billingslea, F. Y. The Bender-Gestalt: an objective scoring method and validating data. *J. clin. Psychol.*, 1948, **4**, 1-27.
2. Cornutt, R. H. The use of the Bender-Gestalt with an alcoholic and non-alcoholic population. *J. clin. Psychol.*, 1953, **19**, 287-290.
3. Gobetz, W. A quantification, standardization, and validation of the Bender-Gestalt test on normal and neurotic adults. *Psychol. Monogr.*, 1953, **67**, No. 6 (Whole No. 356).
4. Pascal, G. R., & Suttell, B. J. *The Bender-Gestalt test.* New York: Grune & Stratton, 1951.
5. Peek, R. M., & Quast, W. *A scoring system of the Bender-Gestalt test.* Hastings, Minn. (Box 292) and Minneapolis, Minn. (2810 42nd St.): Authors, 1951.
6. Robinson, N. M. Bender-Gestalt performances of schizophrenics and paretics. *J. clin. Psychol.*, 1953, **19**, 291-293.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Gellerman's Complex Crossover Design[1]

Julian C. Stanley[2]
Department of Education, University of Wisconsin

In discussing the analysis of his Form I and II Arthur scores, Gellerman [1, p. 129] states: "The *F*-tests were designed after the manner described by Grant . . ." [2]. Yet on page 127 he indicates that assignment to the two counterbalanced order-of-testing groups (I-II, II-I) was not random: "Subjects were selected by the matched-pair method to insure minimal group differences. The factors controlled in selection were sex, chronological age, mental age, previous experience with the Arthur scales, and etiological classification."

Instead of partitioning his total sum of squares into five parts (orders, individuals within orders, sessions, forms, and individuals × sessions within orders), Gellerman could have performed the following more extensive analysis: sexes, pairs within sexes, individuals within pairs, sessions, sessions × sexes, sessions × pairs within sexes, forms, forms × sexes, and forms × pairs within sexes.

Gellerman's data are no longer available, so it is impossible to determine definitely whether or not his conclusions would be altered by t complex breakdown. To illustrate the two d signs and their analyses in considerable detai the present writer divided 28 students random ly into two groups of 14 each. One group r ceived the two forms of a locally prepared fin examination in A-B order; the other grou took Form B first. Form A proved significan ly harder than Form B.

Then Gellerman's design was simulated b *artificially* assigning "sexes" and by matchi like-sexed students across orders. The "erro mean square for form differences dropped fro 11.4 to 4.9. These were *not* Gellerman's scor but it appears reasonable to surmise that h MA difference would also have been fou more significant, rather than less so, if he h employed the longer analysis. No other co parisons were basic to his study.

Gellerman need not have matched subject or else he might well have taken matching e plicitly into account. The basic crossover d sign is much easier to plan and analyze. Ther fore, despite greater possible efficiency of t nine-part analysis, the simpler one seems prefe able unless added comparisons themselves a of interest.

Brief Report
Received March 15, 1954.

References

1. Gellerman, Saul W. Forms I and II of the A thur Performance Scales with mental defectiv *J. consult. Psychol.*, 1952, **16**, 127-131.
2. Grant, David A. The latin square principle i the design and analysis of psychological experi ments. *Psychol. Bull.*, 1948, **45**, 427-442.

---

[1] An extended report of this study may be obtained without charge from Julian Stanley, Education Building, University of Wisconsin, Madison, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4306 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[2] The writer is indebted to V. N. Amble, E. E. Cureton, A. L. Edwards, P. O. Johnson, O. Kempthorne, E. F. Lindquist, and W. J. Moonan for helpful comments concerning an earlier draft of the full report.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Mental Age Equivalents for the Revised Stanford-Binet Vocabulary Test[1]

Edward E. Cureton
University of Tennessee

Neither Terman and Merrill [3] nor McNemar [2] give a table of mental age equivalents for the vocabulary test of the 1937 Revised Stanford-Binet Tests of Intelligence. McNemar presents a table giving the mean and standard deviation of the vocabulary scores for each age group of the standardization sample from 7 to 18 [2, p. 141, Table 44], but says: "We do not deem it necessary or desirable to present a table of mental-age equivalents . . ." [2, p. 140]. This judgment is apparently based on the fact that the vocabulary test, which consists of only 45 items, is not highly reliable.

Both McNemar, and Terman and Merrill note that the vocabulary test has higher correlations with total score than any other. Examiners frequently use it as a brief verbal scale, moreover, so it seems worth while to make available to them the best mental age equivalents that can be derived from the published standardization data.

Derivation of Equivalents

The first column of Table 1 lists the age groups in which the vocabulary test is used in MA determination. The number of cases in each group from 6 to 14 is a little over 200, and in the groups from 15 to 18 it is a little over 100. All children in the standardization groups were within one month, plus or minus, of a birthday, so for practical purposes each age group may be taken to consist of children all of whom are exactly that age.

Column 2 of Table 1, taken from McNemar's Table 44, gives the mean vocabulary scores for age groups 7 to 18. Why he did not give the average for age 6 is not clear, as vocabulary is one of the 6-year tests of Form L. Since these mean scores vary irregularly from their general trend because of sampling errors, is was necessary to smooth them. This was done in two steps. First, a five-point weighted moving average, which preserves linear, parabolic, and cubic trends over each set of five points but removes irregularities of higher order, was applied [1, p. 374, Formula 29.7, $a_0 = 5$]. Since minor irregularities may generate a spurious cubic trend over as many as five points, this formula tends slightly to under smooth, and it leaves unsmoothed the two points at each end of the series. A seven-point formula would have given better smoothing in the middle range, but it would have left three points at each end (half the points in the series) unsmoothed. In the second step, the first

Table 1

Raw and Smoothed Mean Vocabulary Scores and Mean Composite Mental Ages of Age Groups 6 to 18 Inclusive, Revised Stanford-Binet Tests of Intelligence

Age group	Mean vocabulary score		Mean mental age		
	Raw	Smoothed	Raw	Smoothed	Smoothed
6		5.8*	6.09	6.09	6–1
7	7.0	7.0	7.19	7.19	7–2
8	8.3	8.3	8.23	8.29	8–3
9	10.0	9.8	9.43	9.39	9–5
10	11.4	11.6	10.47	10.49	10–6
11	13.7	13.5	11.49	11.54	11–6
12	15.4	15.5	12.39	12.51	12–6
13	17.4	17.1	13.50	13.33	13–4
14	18.2	18.5	13.80	14.01	14–0
15	20.0	19.8	14.73	14.66	14–8
16	21.2	21.0	15.24	15.29	15–3
17	22.0	21.9	15.76	15.77	15–9
18	22.5	22.5	16.09	16.09	16–1

* Extrapolated.

---

[1] Revision of a paper presented to the Tennessee Psychological Association at Nashville, Tennessee, January 23, 1954.

differences of the five-point-smoothed series, including the two unsmoothed points at each end, were set down, and a few additional changes of one digit after the decimal point were made in such a manner as to regularize these differences. From the figures at the top of the smoothed series and of the difference table, one additional value was added to the series by extrapolation at age 6. Column 3 of Table 1 gives the smoothed series.

Terman and Merrill give the Form L and Form M mean IQ's for all age groups of the standardization sample [3, p. 35, Table 5]. The two mean IQ's for each group were averaged and then multiplied by the age, or by the equivalent chronological age divisor in the case of age groups 14 to 18 [3, p. 31, Table 4], to obtain the mean L–M composite mental age. These means are given in Column 4 of Table 1. The same smoothing procedure was applied to them also. In the second stage, no change as great as one digit in the first decimal was made. The smoothed values, in years and hundredths, are given in Column 5 of Table 1. These values were then converted to years and months as shown in Column 6.

The entries in Columns 3 and 6 of Table 1 were then plotted on a large sheet of 1/10-inch, cross-section paper, using 1/2 inch per unit of the vocabulary score and 1/10 inch per month of mental age. It was noted that the trend between the points for ages 10 to 18 was almost exactly linear.

McNemar reported that two subjects, one at age 17 and one at age 18, obtained the highest vocabulary score, 38 [2, p. 140]. This performance was judged by the writer to be approximately equivalent in rarity to IQ 160, or adult mental age 24-0. One point was therefore added to the chart at score 38 and MA 24-0. A straight line drawn through this point and the point for age group 10 passed exactly through the point for age group 18, and no one of the points for age groups 11 to 17 deviated from it by as much as 1/16 inch on the scale used. Hence an upward extrapolation in a straight line from the 10-year to 18-year trend would give exactly the same result as the interpolation to the point at score 38 = MA 24-0. This exact agreement is in considerable part coincidental, of course, but it gives us some further confidence in the accuracy of t
mental age equivalents of scores higher th
that of the 18-year age group average.

The line was continued downward in
smooth curve through the averages for a
groups 6 to 9, and extended one score poi
further downward and two score points fu
ther upward.

The mental age corresponding to eac
score from 5 to 40 was then read from th
curve. Accuracy of reading was checked b
forming a difference table of the mental ag
equivalents, and a few changes of one mont
were made in such a manner as to regulari
these differences. The resulting mental ag
equivalents are given in Table 2.

Table 2

Mental Age Equivalents of Raw Scores on the 19
Revised Stanford-Binet Vocabulary Test

Score	MA	Score	MA*
5	5–4	23	16–4
6	6–3	24	16–10
7	7-2	25	17–4
8	8–0	26	17–10
9	8–10	27	18–4
10	9–7	28	18–11
11	10–2	29	19–5
12	10–9	30	19–11
13	11–3	31	20–5
14	11–9	32	20–11
15	12–3	33	21–5
16	12–9	34	21–11
17	13–3	35	22–5
18	13–9	36	23–0
19	14–3	37	23–6
20	14–9	38	24–0
21	15–3	39	24–6
22	15–10	40	25–0

* Values in this column are interpolated to a sing
point: score 38 (highest made in the standardizati
group; N = 2) taken as equivalent to MA = 24 - 0 (I
160). This interpolation corresponds perfectly to an e
trapolation upward of the curve from scores 12 to 22.

Reliability

Though the mental age equivalents ar
given in years and months to conform to cu
rent practice, the reliability of the test do
not warrant this degree of fineness. No di
rect data on reliability are available, bu
rough estimates may be made. Terman an
Merrill give the median L–M correlation fo
the ages above 6 as .93 [3, p. 47]. They giv
the average correlation between vocabulary an
the L–M composite as .81 [3, p. 302]. If w

raise the .93 by the Spearman-Brown formula, we obtain .964 as the estimated reliability of the L–M composite. Dividing .81 by the square root of .964 gives .825 as an estimate of the correlation between vocabulary and "true" mental age. The lower limit of the reliability of the vocabulary test is the square of .825 or .68. Its reliability could be this low only if it measured exactly the same mental function as the L–M composite, which is clearly not the case. We shall take .75 more or less arbitrarily as an estimate of its reliability in a single age group.

The standard error of measurment is the standard deviation of the scores times the square root of one minus the reliability. Thus $\sqrt{1 - .75} = .5$, so the standard error of measurement is approximately half the standard deviation of the scores. McNemar's Table 44 [2, p. 141] gives the standard deviation of the vocabulary scores for each age group from 7 to 18. Multiplying half of each of these values by the difference between adjacent mental ages in Table 2, at the point corresponding to the entry in Column 6 of Table 1 for the given age group, we obtain the standard errors of measurement in months of mental age. Converting these values to fractions of a year and dividing by the age of each group, or by the equivalent chronological age for age groups 14 and over, we obtain the standard error of measurement of the vocabulary IQ for each age group. The values obtained range from 8.67 to 11.90, with an average of 9.72. We conclude, therefore, that the standard error of measurement of the vocabulary IQ is roughly ten points. This means that in interpreting such IQ's we can place some confidence in the first figure, or the first two figures if the IQ is 100 or over, but none in the last figure. The 95 per cent confidence limits for a given vocabulary IQ are approximately plus or minus 20 points.

Summary

Table 2 gives mental age equivalents of scores on the vocabulary test of the 1937 Revised Stanford-Binet Tests of Intelligence. Any inaccuracies in the table, whether due to the method of its derivation or to the limitations of the original standardization data, are small in comparison to the unreliability of the test.

Despite its limited reliability, which is inevitable in a 45-item test covering the whole range from age 6 to superior adult, this vocabulary test is useful as a short verbal scale.

Vocabulary IQ's may be computed from these mental ages by dividing them by CA up to age 13, and thereafter by the CA equivalents used with the 1937 Revised Stanford-Binet Tests of Intelligence. For the interpretation of any such IQ, its last digit has no significance; the standard error of measurement is about 10 points.

Received January 12, 1954.

References

1. Kendall, M. G. *The advanced theory of statistics*. Vol. 2. London: Griffin, 1948.
2. McNemar, Q. *The revision of the Stanford-Binet Scale*. Boston: Houghton Mifflin, 1942.
3. Terman, L. M., & Merrill, Maud A. *Measuring intelligence*. Boston: Houghton Mifflin, 1937.

NEW BOOKS AND TESTS

Books

Arnold, Magda B., & Gasson, John A. *The human person.* New York: Ronald, 1954. Pp. x + 593. $5.75.

Ten Catholic psychologists met in a workshop "to formulate an integrated theory of personality based on a Christian conception of human nature." This volume reports the planned individual papers and extra chapters on the discussion provoked by them. The material is organized into five parts: theory building, personality structure, personality integration, psychotherapy and self integration, and self integration through religion. The chapter devoted to Frankl's existential analysis is an addition to the American literature on psychotherapy. Logotherapy or emphasis on the client's ideals, aspirations, and deliberate choices deserves detailed systematic treatment. Too often therapeutic theory and practice have neglected a strong emphasis on the client's life as an expression of unique goals with appropriate unique means to accomplish them. The editors present this book as a primary text in personality. It is rather a book on personality theory fairly wide in scope but with relatively little of the empirical findings woven into it. The two chapters on orthodox Christianity and 40 pages devoted to "a research method validating self determination" make it a specialized book rather than a general personality text. This book can serve a wide audience mainly as a point of departure for much needed critical and unfettered discussion of psychology and human values. It raises questions about the relationship between normative and empirical approaches to the human person which all psychologists might well consider. The reviewer wishes it could have been possible for the writers to include a non-Catholic psychologist in their team to increase the objectivity of the presentation and to integrate their views even more with contemporary psychology. — *F. McK.*

Burington, Richard Stevens, & May, Donald Cur Jr. *Handbook of probability and statistics* ca *tables.* Sandusky, O.: Handbook Publishers, 19 Pp. ix + 332. $4.50.

This handbook is prepared in the style of t handbooks which are so important to such fields chemistry and engineering; the binding, format, a materials are most attractive. The text is well s plied with formulas and terse discussions of th meaning and application. The tables are quite co prehensive and include binomial distribution fu tions, Poisson distribution functions, Beta funct ratios, distributions for the familiar statistics, tr onometric functions, natural and common lo rithms, and many other functions. The style a discussions are distinctly technical and the mat matical level of the presentation probably exce the preparation of most psychologists. — *J. R. W.*

Fink, Harold Kenneth. *Mind and performance.* N York: Vantage Press, 1954. Pp. xi + 113. $3.

This is the report of an experimental study learning ability in a number of animal forms. A Arrow Maze, consisting of four parallel pathwa leading off from a neutral area, was employed f all forms. Four successive goals were used for ea of the species, and the criterion for learning was successful trials. Time limits were different for eac species. Human subjects were blindfolded and th maze presented as a stylus maze. Results we treated to yield percentages of the learning abili of man, set at 100. The final rank order of the sp cies tested was man, pig, dog, rat and chick, c turtles (4 species tested), with learning ability the pig about half that of man. Pavlov's concept experimental extinction is employed to explain b havior in learning successive goals. The author di cusses the relationship of his results to Crile's su gestion that behavioral differences may be relat to metabolic functions. — *A. R.*

Freud, Sigmund. *The origins of psychoanalysis; le ters to Wilhelm Fliess, drafts and notes: 188 1902.* New York. Basic Books, 1954. Pp. xi + 486. $6.75.

Freud's letters to Fliess covering the period fro 1887 to 1902 constitute not only one of the most im

---

Note. — The reviews were prepared by the Editor and the Associate Editors, who may be identified by their initials.

ortant publications in the field of psychology in ears, but also a fascinating human document. Never intended for publication, the letters throw a ight on Freud the person, the doctor, the scientist, and the philosopher. At one and the same time they show him as a human being and a genius, a boasting egotist and a disillusioned neurotic, a humanitarian, a devoted family man, and a tyrant with his colleagues. But most of all, they reveal the painful steps by which Freud developed his psychology —his keen observations, his tentative hypotheses, his probing, revising, accepting, and rejecting of oncepts and theories until he was satisfied, and then frequently starting all over again because a new patient provided contradictions not previously yet. Through the letters we follow him in his ups and downs, his satisfactions and disappointments, his self-analysis.

On October 16, 1895 he wrote, "I am practically sure I have solved the riddle of hysteria and obsessional neurosis"; on January 1, 1897, "You can see that I am in the full swing of discovery"; on May 31, 1897 he reports his first inkling of the Oedipus: "It seems as though in sons this death-wish is directed against their father and in daughters against their mother." And so it goes, discovery after discovery told in personal revelations: "I have jotted them down for your eyes alone." His first contact with infantile fantasy in neuroses, the development of libido, the mechanism of regression, the interpretation of dreams, the discovery of everyday psychopathology, of transference, and of many other psychological mechanisms are seen in Freud's first excitement at their unfolding. Included in the volume is Freud's "Project for a Scientific Psychology" which he wrote in 1895 and sent to Fliess for his opinion. This was an attempt to integrate neurology, psychology, and "objective" quantification into a "natural science." This project was later dropped by Freud, and many of its segments adapted in his numerous works. It is nevertheless historically important. A welcome by-product in this publication is Kris's masterful introduction, which throws much light on the letters. Kris's footnotes throughout relate the contents of the letters to Freud's better known works. — *M.K.*

Gill, Merton, Newman, Richard, Redlich, Fredrick C., & Sommers, Margaret. *The initial interview in psychiatric practice.* New York: International Universities Press, 1954. Pp. 423. $6.00. *Records*, 3, 12-in., 33 r.p.m., $4.60.

The first interview with a patient or client embarking on psychotherapy has puzzled experienced and novice therapists alike, and little has been written about it. This volume and its accompanying records present a viewpoint about the initial interview, and provide excellent materials for discussion and instruction. The authors are critical of the traditional psychiatric examination with its rigid structure and probing questions. They also do not entirely accept the position taken by nondirective therapists and some psychoanalysts that the opening interview involves no evaluation or diagnosis. Instead they attempt with considerable success to blend a diagnostic attitude — appraising the patient's need, motivation, and capacity for therapy — with a recognition that the first contact is an integral and often crucial part of the therapeutic process itself. The book consists of 119 pages of exposition and theory with a good review of the literature on the initial interview, and of 278 pages of verbatim illustrative material. Three initial interviews are given in full, with the transcripts on the left-hand pages and a parallel discussion on the right-hand pages. Two of the interviews, by experienced and able therapists, illustrate the application of the authors' approach. These interviews are not perfect, as the authors readily admit, but they are vivid and instructive. The remaining interview, conducted by an inexperienced student, shows most of the mistakes that can be made in interviewing — the interviewer's self-involvement, his inability to really hear the patient, and his defensive dependence on routine questions. The technically excellent phonograph records, available at cost to qualified professional persons and institutions, give the listener many cues that are lacking in the printed transcripts. The authors' moderate and eclectic conception of the initial interview will appeal to most readers as sound and widely applicable. The book and records supply unique and useful tools of instruction for clinical psychology and social work as well as for psychiatry. — *L. F. S.*

Humphreys, Anthony, & Traxler, Arthur E. *Guidance services.* Chicago: Science Research Associates, 1954. Pp. xvii + 438.

This is a text for an introductory course in guidance services. It covers the five subject matter areas reflected in many present courses with appropriately greater emphasis on basic understandings, tools and techniques, and on students' problems than on administration. The authors properly devote a chapter to the future of guidance services, and indicate the trend toward greater professionalization and increase in scientific emphasis. The swell of the trend seems to be definitely in the future if it includes reduction of the nomothetic "do and don't" presentation and an increase in a style which stimulates deeper understanding of the individual student. Humphreys and Traxler should be commended for devoting a quarter of their book to "solving students' major problems." The discussion does not, however, devote much space to students' cases. Student personnel and guidance services are both defined on the same page, but it is not entirely clear to the reader how these overlapping terms differ basically. This whole field of endeavor, as the authors intimate, is undergoing a rapid transition and the pioneer guidance workers in the field who have done much to improve the lot of individual students need to be better linked with the newly and more technically trained worker of the present. Books of this kind can do much to bring the two together. There is still, however, a great gap between the better books in guidance like this and the new volumes on counseling and personnel techniques. — *F. McK.*

Peck, Harris, B., & Bellsmith, Virginia. *Treatment of the delinquent adolescent.* New York: Family Service Association, 1954. Pp. 147 (paper). $2.00.

The authors have given us a realistic picture of the difficulties and successes in the use of psychotherapeutic techniques with delinquent adolescents. There is an excellent appraisal of the limitations and potentialities of group therapy, both with parents and with adolescents themselves. Educators, family and social agencies, and all those working with delinquents will find this book challenging reading. — *B.M.L.*

Warters, Jane. *Techniques of counseling.* New York: McGraw-Hill, 1954. Pp. viii + 384. $4.75.

This is a broad survey of the techniques available for student personnel work. It is written simply and well, and in a manner that will be helpful to the teacher or school counselor in training. The techniques fall largely under the headings of tests, observations, self-reports, status within the group, case study and conference, interview, and group methods. There is no attempt to survey the literature broadly, but the selection is quite appropriate. The reader is given samples of common techniques and an acquaintance with the uses, advantages, and disadvantages of each method. Moreover, a clear presentation of practical issues, especially those with ethical aspects, is included. There is some discussion of counseling, but the integration of several techniques into actual counseling situations is left for advanced clinical training. Miss Warters faces the dilemma of most writers in the field: how to encourage school personnel to use more objective techniques and at the same time instill sufficient caution concerning the limitations in the use of the techniques without supervised laboratory or clinical training. This problem is handled rather well. The reviewer, however, would have favored stronger and repeated emphasis on the necessity for more than verbal acquaintance with the techniques. — *F. McK.*

Wolberg, Lewis R. *The technique of psychotherapy.* New York: Grune & Stratton, 1954. Pp. xiv + 869. $14.75.

Good psychotherapy, says Dr. Wolberg in his preface, "must depend on a basically sound therapeutic structure, irrespective of the specified 'kind' of psychotherapy done and without reference to the individual 'style' of the therapist. The purpose of this book is to delineate such a structure." How well the book is judged to achieve this ambitious aim will depend on the experience and therapeutic predilections of those who read it. Few will deny, however, that this encyclopedic volume represents some of the most practical and down-to-earth writi about psychotherapy yet to appear in the literatu Rich in excerpts from recorded interviews and ented around recurrent and realistic problems therapeutic practice, it often has the flavor o fruitful supervisory session. If its eclecticism relative naiveté with respect to theory someti lead to contradictions and rather too easy soluti to difficult problems in the shadowland of th peutic knowledge, the book nevertheless presen large distillation of practical wisdom won fr long and intensive experience. It makes possible intimate acquaintance with one thoughtful and w grounded point of view toward therapeutic cond One regrettable shortcoming, constituting a ch lenge to psychologists, is that in spite of its frien liness to psychology among the professions offeri needed psychotherapeutic services, and its recogniti of some of the contributions to therapeutic thou made by psychologists, this big book makes very li reference to the insightful investigative work d by psychologists or, indeed, to research in the f at all. As a summary of one therapist's conside wisdom, *The Technique of Psychotherapy* is comp hensive and admirable. For secure knowledge, shall not surprisingly have to wait. — *E. J. S.*

Wolstein, Benjamin. *Transference: its meaning function.* New York: Grune & Stratton, 1954. xiii + 206. $5.00.

Written with sensitivity, scholarship, and log sophistication, this careful historical analysis of concept of transference represents a genuine con bution to psychoanalytic theory. In the tradition sociated with Harry Stack Sullivan, the develops the thesis that psychoanalysis is fun mentally the study of interpersonal processes, the therapeutic utility of transference phenom lies in their enabling both therapist and patient arrive at greater insights into the latter's modes relating to his fellows, and that the transfere situation is not only a medium by which already tablished values may be re-examined but the se bed of new values and new orientations toward h man relationships. In the course of this developme Dr. Wolstein demonstrates that he has ideas, wri with style and a gentle persuasiveness, and co petently levies on a quiet but impressive learni It is a disappointment, however, to find a psycho gist writing so authoritatively in this field with total disregard for the bulk of pioneering resea that has been the central contribution of his disc line to therapeutic knowledge. An integration empirical findings with Dr. Wolstein's inform level of theorizing could be a vast boon to psych therapeutic thought and practice. — *E. J. S.*

Woodworth, Robert S., & Schlosberg, Harold. *Ex perimental psychology.* (Rev. Ed.) New Yor Holt, 1954. Pp. xi + 948. $8.95.

The classic *Experimental Psychology*, first pub lished in 1938, has been given a thorough revisi Most chapters are rewritten, some are new, and

e brought up to date. Over half of the 2,480 en-
ies in the bibliography were not cited in the earlier
lition. In addition to its values for general psy-
ology, what kind of a foundation does this book
ovide for students who will specialize in clinical
sychology? An appraisal from this limited view-
oint is generally favorable. The three chapters on
notion have been revised extensively, with level of
ctivation as an organizing concept. Learning is
eated so as to provide a good base for the applica-
ons of learning theory to personality development
nd psychotherapy. There is a new chapter on mo-
vation. Dynamic aspects of perception and memory
ave been given increased recognition. Of more im-
ortance than these details, however, is the general
mosphere of the volume. Its consistently functional
nd objective attitude, its persuasive demonstrations
f the values of experimental method, and its clear
nd interesting style of writing will lead students
ot only to learn experimental psychology but also
appreciate and enjoy it. — *L. F. S.*

Vulfeck, Joseph W., & Bennett, Edward M. *The language of dynamic psychology as related to motivation research.* New York: McGraw-Hill, 1954. Pp. 111. $4.00.

This volume is a glossary of terms commonly ap-
earing in the field of motivation and apt to be
neaningless or confusing to the lay reader. Each of
he terms presented is accompanied by a definition
urnished by Wulfeck and by some explanatory
omment by Bennett. The volume is sponsored by
he Advertising Research Foundation and is de-
igned to "help the advertising man to keep his
erminology and his understanding abreast of cur-
ent activities in the field of motivation research."
To the best of this reviewer's empathic judgment
he book would seem to serve its purpose. The pro-
essional psychologist, however, will find little here
enlighten him. In fact, he may occasionally be
arred by expressions such as this, "Instinct (Psy-
hiat): a wastebasket category for the origin of any
ehavior which is not learned." — *W. A. H.*

Books Received

Barrell, Joseph. *A philosophical study of the human mind.* New York: Philosophical Library, 1954. Pp. xii + 575. $6.00.

Beck, Samuel J. *The six schizophrenias.* Res. Monogr. No. 6. New York: American Orthopsychiatric Association, 1954. Pp. xiii + 238. $5.00.

Bischof, L. J. *Intelligence.* Doubleday Papers in Psychology, No. 5. Garden City, N. Y.: Doubleday, 1954. Pp. vi + 33 (paper). 85¢.

Braatøy, Trygve. *Fundamentals of psychoanalytic technique.* New York: Wiley, 1954. Pp. xi + 404. $6.00.

Campbell, Paul, & Howard, Peter. *Remaking men.* New York: Arrowhead Books, 1954. Pp. 126. $1.50.

Cronbach, Lee J. *Educational psychology.* New York: Harcourt, Brace, 1954. Pp. xxvii + 628. $7.50.

English, O. Spurgeon, & Finch, Stuart M. *Introduction to psychiatry.* New York: Norton, 1954. Pp. viii + 621. $7.00.

Frank, Lawrence K., Harrison, Ross, Hellersberg, Elizabeth, Machover, Karen, & Steiner, Meta. *Personality development in adolescent girls.* Monogr. Soc. Res. Child Develpm., 1951, 16, No. 53. New Orleans, La.: Child Development Publications, 1953. Pp. 316 (paper).

Frosch, John. (Ed.) *Annual survey of psychoanalysis.* Vol. II, 1951. New York: International Universities Press, 1954. Pp. xiv + 724. $10.00.

Fuller, John L. *Nature and nurture.* Doubleday Papers in Psychology, No. 4. Garden City, N. Y.: Doubleday, 1954. Pp. vii + 40 (paper). 85¢.

Haber, Heinz. *The physical environment of the flyer.* Maxwell AFB, Ala.: Air University, 1954. Pp. vi + 179.

Himes, Joseph S. *Social planning in America.* Doubleday Short Studies in Sociology. Garden City, N. Y.: Doubleday, 1954. Pp. ix + 59 (paper). 95¢.

Hodgson, Kenneth W. *The deaf and their problems.* New York: Philosophical Library, 1954. Pp. xx + 364. $6.00.

Ittelson, William H., & Cantril, Hadley. *Perception.* Doubleday Papers in Psychology, No. 7. Garden City, N. Y.: Doubleday, 1954. Pp. ix + 33 (paper). 85¢.

Joseph, Harry, & Zern, Gordon. *The emotional problems of children.* New York: Crown Publishers, 1954. Pp. ix + 310. $3.75.

Katz, Daniel, Cartwright, Dorwin, Eldersveld, Samuel, & Lee, Alfred M. (Eds.) *Public opinion and propaganda.* New York: Dryden Press, 1954. Pp. xx + 797. $6.25.

Lazarsfeld, Paul F. (Ed.) *Mathematical thinking in the social sciences.* Glencoe, Ill.: Free Press, 1954. Pp. 444. $10.00.

Leonard, Edith M., Van Deman, Dorothy D., & Miles, Lillian E. *Counseling with parents.* New York: Macmillan, 1954. Pp. vii + 330. $3.75.

Odlum, Doris M. *Psychology, the nurse and the patient.* (2nd Ed.) New York: Philosophical Library, 1954. Pp. 168. $4.75.

Pearson, Gerald H. J. *Psychoanalysis and the education of the child.* New York: Norton, 1954. Pp. x + 357. $5.00.

Piaget, Jean. *The construction of reality in the child.* New York: Basic Books, 1954. Pp. xiii + 386. $6.00.

Pickford, R. W. *The analysis of an obsessional.* New York: Norton: 1954. Pp. xii + 234. $4.00.

Sarason, Seymour B. *The clinical interaction.* New York: Harper, 1954. Pp. x + 425. $5.00.

Savage, Leonard J. *The foundations of statistics.* New York: Wiley, 1954. Pp. xv + 294. $6.00.

Schanck, Richard L. *The permanent revolution in science.* New York: Philosophical Library, 1954. Pp. xvi + 112. $3.00.

Senn, Milton J. E. (Ed.) *Problems of infancy and childhood, 1953.* New York: Josiah Macy Jr. Foundation, 1954. Pp. 196. $2.75.

Solomon, Joseph C. *A synthesis of human behavior:* New York: Grune & Stratton, 1954. Pp. xii + 265. $5.50.

Sullivan, Harry Stack. *The psychiatric interview.* New York: Norton, 1954. Pp. xxiii + 264. $4

Taylor, W. S. *Dynamic and abnormal psychol* New York: American Book Co., 1954. Pp. xiv 658.

Watson, Robert I. *Psychology as a profess* Doubleday Papers in Psychology, No. 8. Gar City, N. Y.: Doubleday, 1954. Pp. x + 65. 85

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

Thematic Apperception Test: A Tentative Appraisal of Some "Signs" of Anxiety

Gardner Lindzey and Arthur S. Newburg
Harvard University[1]

Generalizations concerning the relation between aspects of test performance and those attributes of the empirical world that the test intends to predict or diagnose stem from many sources. Perhaps the most fertile of these is careful observation by sensitive and experienced clinicians. By contrast, the role of controlled, empirical study in creating such generalizations is relatively slight. When it comes to evaluating or instituting a system of checks, however, the contribution of controlled investigation is very great. The free-ranging sensitivity of the clinical observer, which makes him an ideal generator of new ideas, also promises that individual bias and autistic factors will play some role in determining his statements. Thus, only when the clinician's generalizations have been submitted to test under circumstances designed to eliminate observer bias can we place much enduring confidence in them.

In the present study we selected a small number of generalizations concerning the Thematic Apperception Test and its sensitivity to anxiety and attempted to test these under some degree of empirical control. As a result of an earlier survey of the literature [3], we had available over 500 statements relating aspects of TAT response to characteristics of the storyteller. From this list we selected 18 generalizations concerning anxiety and attempted to translate each of these into an objective scoring system that would permit us to score TAT protocols reliably. In addition, we sought to secure a reasonable, independent measure of anxiety that could be correlated with our TAT "signs."

Procedure

Subjects

The subjects of this study were 20 undergraduate males who had volunteered from an introductory course in psychology at Harvard College to participate in a program of personality study. They were paid for their time on an hourly basis at customary student rates. The subjects were selected so as to be heterogeneous in regard to such factors as socioeconomic and ethnic background, academic performance, and extracurricular activities.

Thematic Apperception Test

Administration. The TAT was administered under standard conditions with the two sessions separated by 24 hours or more. In all cases the stories were electrically recorded without the subjects' knowledge and later transcribed verbatim.

Scoring. Our scoring procedures were based on statements culled from the literature. In describing these we will first present the pertinent quotation, and following this we will summarize the variable or variables that were designed to correspond to those mentioned in the statement. The reader will notice that we have not attempted to exhaust the meaning of these quotations but have simply selected, somewhat arbitrarily, certain variables that seemed especially pertinent or easy to objectify. In all there were 18 signs or variables, presumably sensitive to anxiety, that we attempted to measure.

While scoring, the examiner knew only that the subjects were male and college students. Every variable was scored individually for each story, and all stories told to a given card were scored before proceeding to the next set of stories. With the exception of the second variable, all scores were dichotomous as the scorer simply indicated for each story that the variable was present or not present (high or low). In some cases, this required setting relatively arbitrary cutting points. Thus, for all but the second variable, the subject's score was made up of the sum of 20 yes and no scores and consequently could range from 0 to 20.

---

[1] This study is part of a program of research conducted at the Harvard Psychological Clinic under the direction of Professor Henry A. Murray. The research is supported by grants from the Rockefeller Foundation and the Laboratory of Social Relations, Harvard University. We are grateful to Shirley Shapiro for her assistance in the statistical analysis.

> [Anxiety is indicated by the] . . . presence of disruptions in the consistency of form characteristics, especially in organizational elements, in vacillation from concept to picture-dominated responses, and in rejections and refusals; content characteristics . . . of overt aggression, depression, mental conflicts or other outstanding emotional states . . . [Henry, 2, p. 48].

1. *Intrapicture vacillation from concept- to picture-dominated responses.* A concept-dominated response is one whose main theme deals with an idea the picture suggests rather than the picture itself. It may be abstract, symbolic, or allegorical, and often relates the events of the story to the general case. A picture-dominated response is one whose main theme deals directly with the events of the picture itself. Often past and future action is omitted, and in general the story action takes place in the here and now of the picture. At its most restricted level the response may be only a description of the picture with little or no plot. If during a single story there was one or more shifts from one type of response to the other, the story was scored "vacillation."

2. *Interpicture vacillation from concept- to picture-dominated responses.* Each story was scored as concept, picture, or vacillation as defined for the preceding variable. Then the number of changes from concept to picture domination and vice versa in the 20 stories was counted.

3. *Rejections and refusals.* The subject makes a negative comment about the picture, or his ability to create a story from the picture. He contradicts or denies or reverses part of his story. He refuses to tell or finish a story.

4. *Outstanding emotional states.* The characters show strong and persistent overt aggression, depression, mental conflict, fear, extreme happiness, love, worry, etc.

> . . . Phantasies with marked anxiety were "characterized by moving, dramatic situations and intense, comparatively clearcut conflicts" [p. 75]. . . . A high incidence of verbs denotes a kinetic release in the phantasy of anxious tensions in the narrator [p. 77]. . . . Total number of verbs/total number of adjectives. High values connote restless, forceful, dramatic action in the phantasies, expressing libidinal tensions and anxiety in the subject [p. 79]. . . . the phantasies in an *anxiety state* are brief; the action is most dramatic (highest verb-adjective quotient) and often compulsive; alternatives of conation are most frequently sough special expressions connoting vagueness, he tation, and trepidation are freely used; a direct identification of the narrator with ch acters in his phantasy frequently occur . phantasied situations are left as unresolved the underlying emotional conflicts of the s ject [pp. 80-81] [Balken and Masserman, 1

5. *Intense conflicts.* The conflict may be intra- interpersonal. It should involve violent emotio states and if it is intrapersonal the individual shou be seen as "torn" by the two sides of the conflict.

6. *Vagueness and hesitation.* The subject mak *two or more* statements showing either vaguene or hesitation or both, e.g., "I'm not sure," "I do know," "I can't tell."

7. *Self-identification.* The narrator openly adm that one of the characters resembles or acts like hi self. He indicates that part or all of the story taken from his own life.

8. *Unresolved conflicts.* The story has no o come. Conflicts, situations, plots are left unfinish Statements like "I don't know how this comes o "I don't know what comes after this" are comm Or, conflict situations are seen as remaining so, e "And he never forgives her for that," "And t couple continues to quarrel."

9. *Interjections.* The subject makes *two or m* statements not directly related to the story he telling. These may be in the form of comments ab the picture: "My God, what's that?" "I'll never able to get anything out of that." Or he may br into his story to ask the experimenter a questi "Is this the kind of story you want?" "Haven' seen this picture before somewhere?" Or he m break the continuity of his story by suddenly co menting on his style, delivery, plot, etc.: "This quite an allegory, isn't it?"

10. *Verb/adjective quotient.* First the total nu ber of verb forms was counted. The whole predic was counted as one verb. Then the total of ver was divided by the total number of adjectives and ratio of 3.5 or better was considered high and the f other ratios low. With this variable, and the fo lowing variable, we used a random number ta to select three stories from each set of protocols a used these stories as unbiased estimates of the tual quotient for the entire 20 stories.

11. *Number of adjectives per 100 words.* Th number of adjectives was computed in the sa manner as in the previous variable. This figure w divided by the total number of words and a ratio .12 or greater scored high and all ratios less th this scored low. Again the sampling technique d scribed above was used.

12. *Briefness.* The total number of words in th story was counted. Any story with less than 1 words was classed as brief and all others were co sidered long.

If the testing situation itself significantly rais the level of anxiety, a not uncommon sequel

> invariance of stories on the level of object description: "The violin looks like a Stradivarius" [pp. 57-58]. Wishes may also be reactivated by memory. Individuals who suffer grief, shame or anxiety may tell such stories. . . . In our experience the sequence of memory-wish has usually signified serious disturbance of the inner life [p. 64]. Only when a protocol is invariant with respect to the level on which the stories proceed may we assume that this level is a literal representation of the predominant level on which the individual functions. An example of such invariance is the individual whose anxiety is reflected in his stories by exclusive reference to the level of feeling and expectations [p. 103] [Tomkins, 7].

13. *Invariance on the level of object description.* With the exception of opening and concluding sentences such as " 'Lemme' see," "Well, I guess that's about it," every scorable sentence must be restricted to the simple, physical description of the objects or people in the picture. Neither the feelings nor thoughts of any of the people in the story may be present in any sentence.

14. *Invariance on the level of feeling and expectation.* With the exception of the beginning and ending phrases noted in the previous variable, every sentence must contain affect. It is not necessary that the kind or quality of emotion expressed be the same. The narrator must attribute emotion or expectation throughout the story. "I feel that...," "He is worried (afraid, in love, fears, hates, longs for, etc.)." A single sentence in which this was absent scored the story as negative.

15. *The sequence of memory followed by a wish.* Either the narrator or one of the characters recalls material from the past and this stimulates him to express a desire for something now. A sentence, clause, or phrase containing a verb of remembering is followed by a passage containing a verb of wishing or desiring, e.g., "The old man recalls his wife fondly; he wishes she were still alive."

> No suggestive diagnostic features other than sporadic blocking, flurries of anxiety in the course of telling the stories, and frequent themes of apprehensiveness are to be expected; even these do not occur frequently [Schafer, 6, p. 46].

16. *Blocking.* The subject's response to the picture is blocked; he is incapable of responding quickly and freely to the stimulus. A long pause before he responds, or two or more pauses during the story, indicates this. Any statement he makes which has a content indicating that he feels unable to produce a story, or that producing one would be difficult, was scored. Statements such as "I can't get any further," or "I seem to be stuck," are examples.

17. *Apprehensiveness.* The presence of themes in which there is an anticipation of future evil or misfortune. The narrator or one of the characters in the story has feelings of dread or foreboding about the future. He has an expectation that unpleasantness lies ahead. "Things don't look too bright for the future." "I'm afraid the situation gets worse for these people."

> Anxiety neurotics had many plots emphasizing sudden physical accidents and mental traumas such as loss of wife, mother, sweetheart or jobs, house burning down, stock market crash, etc., thus reflecting their own fears and the instability of their world [Rotter, 5, p. 30].

18. *Trauma.* The content of the story deals with a theme in which: (*a*) one of the characters has a sudden physical accident, e.g., "and then he fell and broke his leg," " . . . when suddenly a rock dropped on her"; (*b*) someone undergoes a severe mental trauma, characterized by its lasting effects.

Reliability. In estimating the reliability of the TAT scoring we selected at random 100 stories, and a reliability rater scored these for each of the variables. These ratings agreed with the experimental ratings in 82% of the cases, indicating a relatively high degree of association between the two sets of ratings.

Criterion Measure

The primary independent measure was a "diagnostic council" rating based upon a very wide variety of information derived from observation, self-report, situational tests, and a large number of indirect measures. After preliminary analyses, this information was presented individually for each subject and discussed in a council of clinical psychologists. During this session the group assigned a tentative rating to each subject to indicate the importance of anxiety as a determinant of his behavior. Later all the subjects were considered jointly and ranked in terms of the degree or importance of anxiety. In addition to the diagnostic council ratings, scores were available for the Psychosomatic Inventory which presumably should bear some relation to anxiety. We obtained a rank-order correlation of .69 between the total score of the inventory and the clinical ranking.

Statistical Analysis

In most of our analyses we relied upon rank-order correlation as a statistic suited to the relative crudeness of our data. In attempting to relate our "signs" to the subscores of the Psychosomatic Inventory, we used a technique described by Mosteller [4] for estimating correlation coefficients. This short-cut method was used as these relations were of relatively slight

Table 1
Relation of TAT "Signs" of Anxiety to Criterion Measures

TAT variable	Clinical ratings	Psychosomatic inventory					
		Total score	Psycho-somatic	Gastro-intestinal	Nervous tension	Worry	Fear
1. Intrapicture vacillation	.05†	−.32†	−38 ‡	−49 ‡	−21 ‡	−28.5‡	− 1
2. Interpicture vacillation	−.38	−.04	−14	−19.5	− 8.5	+ 5.5	− 4.5
3. Rejection and refusal	.30	.40*	+33.5	+13.5	+26.5	+27.5	+ 3
4. Outstanding emotional states	−.06	−.02	+10	− 1.5	− 1	−37.5	− 7
5. Intense conflicts	.03	.06	+ 8.5	+ 3.5	+ 9.5	−30	−10.5
6. Vagueness	.40*	.15	+ 7.5	− 7.5	−10	+29.5	+18.5
7. Direct identification	.39*	.09	− 1.5	− 5	+ 7	− 1.5	0
8. Conflict unresolved	.22	−.02	− 5.5	−28	− 1	−29	− 1.5
9. Interjections	.06	.21	+17.5	+ 9	+25.5	+ 9.5	+11
0. Verb/adjective quotient	−.09	−.13	− 9	−31.5	−23	+ 2	− 8
11. Adjectives per 100 words	.13	.09	+ 7.5	+33	+11.5	− 5	− 1
12. Briefness	.15	.31	+36	+51	+10	+26	+ 9
13. Object description	.38*	.53*	+43	+ 2	+37.5	+32.5	+35
14. Feeling and expectation description	.22	.02	+19.5	+ 3	− 1.5	− 9	−22.5
15. Sequence: memory to wish	−.16	−.09	−11	−22.5	−21.5	−29.5	−18.5
16. Blocking	.33	.20	+20.5	+11	+ 5.5	+ 2.5	+ 1.5
17. Apprehensiveness	.05	−.04	− 8	− 1	− 2	−33	−11.5
18. Trauma	−.19	.10	+ 7.5	+22	+23	+ 1.5	− 4

* Significant at 5% level with one-tailed test.
† Rank-order correlation.
‡ Mosteller's correlation approximation.

interest, and indicating the direction of the relation and the existence of strong relationships seemed sufficient. A perfect correlation would be signified by a score of 96 using this technique.

Results and Discussion

Our major results are summarized in Table 1 where we find that, of the 18 TAT signs, 13 show some trend toward positive association with the clinical rating of anxiety. Three of the positive correlations (vagueness and hesitation, self-identification, object description) are significant at the 5% level, and there is one negative correlation (interpicture vacillation from concept- to picture-dominated response) of approximately the same magnitude. The findings are not encouraging so far as t utility of the particular TAT measures e ployed here are concerned. Although there some evidence of a tendency toward associatio between the TAT variables and our criterio measure, as witnessed by the preponderance positive correlations, this seems a very tentati link, and the existence of relatively high neg tive correlations suggests dramatically how f we are from an adequate formulation of ho to measure anxiety with this instrument.

The correlations between the TAT sig and the Psychosomatic Inventory are also sum marized in Table 1. In general, they confor

Table 2

Rho Intercorrelation of TAT "Signs" of Anxiety

TAT variable	Intrapicture vacillation	Interpicture vacillation	Rejections and refusals	Emotional states	Intense conflicts	Vagueness and hesitation	Self-identification	Unresolved conflicts	Interjections	Verb/adjective quotient	Number of adjectives	Briefness	Object description	Feeling and expectation	Memory followed by wish	Blocking	Apprehensiveness
Intrapicture vacillation																	
Interpicture vacillation	–.27*																
Rejections and refusals	–.28	–.10															
Emotional states	.25	–.03	–.18														
Intense conflicts	.25	–.25	–.13	.85													
Vagueness and hesitation	.14	–.52	.53	–.41	–.25												
Self-identification	.25	–.20	.52	–.19	.00	.33											
Unresolved conflicts	.60	–.20	–.03	.41	.53	–.04	.38										
Interjections	–.04	–.20	.75	.03	.21	.56	.19	–.01									
Verb/adjective quotient	–.18	.21	–.40	.36	.25	–.30	–.36	–.05	–.27								
Number of adjectives	.25	–.11	.37	–.36	–.33	.21	.42	–.04	.16	–.89							
Briefness	–.71	.26	.35	–.37	–.50	.02	–.12	–.33	–.02	–.02	.10						
Object description	–.10	.26	.54	–.32	–.37	.18	.47	.02	.11	–.30	.32	.30					
Feeling and expectation	–.26	.06	.21	–.11	–.15	.08	.12	.15	.03	–.24	.12	.41	.14				
Memory followed by wish	–.03	–.06	.03	.39	.34	–.03	–.17	.16	.37	.08	–.28	–.09	–.09	.25			
Blocking	.07	–.48	.75	–.29	.09	.78	.32	.16	.59	–.27	.18	.18	.20	.10	–.10		
Apprehensiveness	.19	–.05	–.18	.96	.83	–.31	–.22	.38	.13	.31	–.30	–.37	–.43	–.12	.36	–.07	
Trauma	–.02	.23	–.12	.63	.45	–.39	–.28	.00	.17	.12	.05	–.11	–.12	–.39	.17	–.17	.69

* A rho of .44 is required for significance at the 5% level.

to the tendencies revealed in the clinical rating correlations. Especially for the total score and the psychosomatic subscore the direction of the correlations is quite similar to those we have already discussed, although the magnitude is generally less. In view of the presumed lesser sensitivity of this paper-and-pencil instrument, these findings are to be expected.

In Table 2 we present the intercorrelations between the TAT signs. If we exclude the first two variables, which appear to have been reversed in their scoring, the general tendency is for low positive correlations. In general this table of intercorrelations does not suggest a set of indices for the same or very similar variables. If some of these signs are in fact highly indicative of anxiety, then it seems evident that others are *not* very sensitive to this variable. Nevertheless, several interesting clusters appear. In one we find the following variables: strong emotional states, intense conflicts, apprehensiveness, and trauma, while in the other we find: rejections, hesitations and vagueness, interjections, and blocking. The question of how much these represent related psychological variables and how much the association is artificially produced by scoring the same story attributes under multiple headings is not clearly answered by our data. Certainly an examination of the definitions of these eight variables suggests that some of these intercorrelations are the results of multiple scoring of the same aspects of the stories.

Speculation concerning the relatively strong negative correlation between our measure of interpicture vacillation and our criterion measure of anxiety suggested that a variable of rigidity or inflexibilty might be mediating this relationship. A low picture-vacillation score implies that the subject adopts an initial set toward the story construction process and maintains this without variation throughout the series of pictures. The widely observed "freezing" effect of anxiety makes it understandable that the individual high in anxiety might be relatively invariant in his approach to storytelling. Given this initial hypothesis, we looked for other variables that could be seen as related in some way to rigidity, lack of spontaneity, perseverative tendencies, or inflexibility. We then set up new ranks based upon the pooled scores from these variables. First we combined scores from the interpicture vacillation and intrapicture vacillation variables, as these measures seemed obviously to be getting at related material in spite of their low intercorrelation. For these two variables we reversed the direction of our original scoring because of the observed negative correlation and our *post hoc* rationale. The resulting correlation (.47) was somewhat better than the .38 correlation that our single measure had given us. Next we included scores for the tendency to simply describe objects or aspects of the stimulus picture. This combined rank showed a correlation of .56 with our criterion. Following this we developed a rank order based on these three variables plus scores for the tendency of the storyteller to specifically identify part of the story as coming from his own life or one of the characters as resembling himself. This last variable suggests a concreteness and lack of spontaneity that goes with the other variables we have already mentioned. The rank order based upon these variables correlated with our criterion measure .58. Finally, we introduced scores for vagueness or hesitation of the storyteller, reasoning that this indecisiveness was related to the lack of spontaneity we have already commented upon. This ranking showed a .63 correlation with the criterion measure.

The fact that through the use of three variables we were able to secure a correlation of .56 with our criterion and that with the addition of two more variables we could raise this to .63 suggests that potentially the TAT may be a relatively sensitive indicator of anxiety. On the other hand, in evaluating these findings it is important to consider the small size of our sample and its composition. Our subjects all met some criteria of normality although at least four of them presented rather serious adjustment problems. Even these four, when compared to severely disturbed patients, probably could be characterized as only moderately anxious. Thus our results are based upon a restricted range of the variable of interest.

The relative success of our formal measures in predicting anxiety when compared to the content measures suggests that analysts of the TAT might well devote more of their time to this type of variable. It is, of course, possible that this finding may be rather specifically linked to the nature of the variable under study.

In summary, our results are somewhat discouraging for the utility of common generalizations concerning anxiety and the TAT. Nevertheless the fact that we were able to find relatively substantial correlations between our criterion measure of anxiety and small clusters of TAT variables that bore some rational relation to each other suggests the potential utility of the instrument in this area. If we take seriously the highly tentative evidence of this study, it seems that the TAT of the anxious person is characterized by an excessive sameness or rigidity in the approach of the storyteller, a preference for limiting himself to simple object description, a readiness to relate parts or characters in the stories explicity to himself, a tendency to be vague or hesitant in presenting his stories, and also a readiness to reject his productions and refuse to tell or complete his stories.

Received April 1, 1954.

References

1. Balken, Eva R., & Masserman, J. H. The language of phantasy: III. The language of the phantasies of patients with conversion hysteria, anxiety state, and obsessive-compulsive neuroses. *J. Psychol.*, 1940, **10**, 75-86.
2. Henry, W. E. The analysis of fantasy: the the-

matic apperception technique in the study of behavior. Unpublished manuscript, Univer. of Chicago, 1950.

3. Lindzey, G., Tejessy, Charlotte, Davids, A., & Heinemann, Shirley H. Thematic Apperception Test: a summary of empirical generalizations. Unpublished manuscript, Harvard Univer., 1953.
4. Mosteller, F. On some useful "inefficient" statistics. *Ann. math. Statist.*, 1946, **17**, 377-408.
5. Rotter, J. B. Studies in the use and validity of the Thematic Apperception Test with mentally disordered patients. I. Method of analysis and clinical problems. *Charact. & Pers.*, 1940, **9**, 18-34.
6. Schafer, R. *The clinical application of psychological tests: diagnostic summaries and case studies.* New York: International Universities Press, 1948.
7. Tomkins, S. S. *The Thematic Apperception Test.* New York: Grune & Stratton, 1947.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

Brief Intellectual Assessment of Patients with Behavior Disorders[1, 2]

Herbert W. Eber, Carl M. Cochrane
University of North Carolina
and Albert A. Branca
VA Hospital, Fayetteville, North Carolina

In treatment-oriented situations differentiation between *essentially normal* and *very low* intelligence is often sufficient. The Wechsler-Bellevue requires lengthy administration time disproportionate to the information required in these cases. The SRA Non-Verbal Form [1], a ten-minute test (plus one minute for scoring) with well-standardized norms and a reliability of .85, does not sample intelligent behavior broadly enough to confirm *very low* intelligence; in this study, it was evaluated for its ability to rule out *very low* intelligence.

As the use of a normal population would have spuriously raised the validity of the technique, the SRA was here evaluated in the setting for which it is recommended. The subjects of the experiment were all the patients without brain damage, seen during a six-month interval in the NP service of this facility, in whose cases the ruling out of very low intelligence was a real problem. This group of 32 patients was modal around the point of differentiation, and represented a real challenge to the technique. The mean Wechsler IQ of the group was 86.9 and the SD was 17.6. Both the Wechsler (Form I or II) and SRA were administered to each patient individually.

For purposes of comparison the scores on both tests were converted into percentiles. These were classified according to the method Wechsler used in his statistical definition of intellectual levels. His cutting point between "borderline" and "dull normal" was set at minus 2 PE, the 8.9th percentile. In this study, the class *essentially normal* was defined as at or above the 9th percentile; *very low* was defined as below this point.

The results are shown in the following contingency table:

	SRA Low	SRA Normal
Wechsler normal	7	13
Wechsler low	12	0

Thus it was shown that any patient scoring at the 9th percentile or higher on the SRA scores above this mark on the Wechsler. The converse relationship does not hold. The correlation (product-moment of standard scores derived from percentiles) is .75; but the predictive value for our purpose is greater than the correlation shows, since an *essentially normal* score on the SRA apparently assures a like score on the Wechsler. Thus the SRA may be regarded as one-sided instrument, capable of ruling out *very low* intelligence but not capable of confirming it.

Conclusion. On the basis of the 32 cases, it was concluded that the SRA Non-Verbal Form was a useful screening device for patients with behavior disorders when the only task is to select those high enough in intelligence to be considered *essentially normal*. Patients without brain damage who score at or above the 9th percentile on the SRA may be considered of normal intelligence, while those scoring below this point need further evaluation to obtain a broader based estimate of intellectual level.

Brief Report
Received June 25, 1954.

Reference

1. McMurry, R. N. & King, J. E. *SRA Non-Verbal Form.* Chicago: Science Research Associates, 1947.

---

[1] From the Veterans Administration Hospital, Fayetteville, North Carolina.

[2] An extended report of this study may be obtained without charge from Albert A. Branca, Chief, Clinical Psychology, VA Hospital, Fayetteville, North Carolina, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4344 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

Detection of Suicidal Patients: An Example of Some Limitations in the Prediction of Infrequent Events

Albert Rosen
VA Hospital, Minneapolis, Minnesota
and the University of Minnesota[1]

Since the detection of suicidal patients is an important responsibility of clinical psychologists, a number of investigators have proposed or evaluated certain signs, configurations, and items from psychological tests for their identification [1, pp. 290, 326; 7; 8; 9; 12; 14; 15; 18; 19; 21]. The purposes of this paper are (*a*) to suggest modifications in the classification of patients in suicide research, (*b*) to emphasize some of the limitations inherent in the prediction of suicide, and (*c*) to indicate the general applicability of such limitations in the prediction of any other behavior or event of infrequent occurrence.

Classification of Patients in Suicide Research

The term "suicidal" has been used to describe patients demonstrating heterogeneous kinds of behavior, such as (*a*) suicide thoughts, (*b*) suicide threats, (*c*) suicide attempts, and (*d*) severe depression. In previous research on suicide predictors, a common practice has been the combining of test data from such individuals with data from those who actually committed suicide.

Differentation of "Suicidal" Subgroups

The need for more precise classification in suicide research has been indicated by Farberow [6] and by Rosen, Hales, and Simon [16], who demonstrated that "suicidal" subgroups can be differentiated by means of psychological tests. The two studies closely agreed in indicating that, as a group, patients who express suicide thoughts or suicide threats are much more severely disturbed than either patients who have made a suicide attempt or patients-in-general, and that patients in the latter two categories are similar to each other.

Suicide Rates

Additional evidence of the need for greater refinement in the classification of patients in suicide research is provided by statistics on the incidence of suicide. Suicide is an extremely infrequent event, and very few of those patients who are severely depressed or who express suicide thoughts or threats actually commit suicide. Thus, data from such patients should not arbitrarily be grouped under the class term "suicidal." Examination of the following data on suicide rates provides substantiation of these statements.

Suicide rate for the general population. In the United States, the suicide rate for the general population is 11.4 in 100,000 [20, p. 75]. Reliable statistics for the frequency of suicide *attempts* in the general population are unavailable because of the difficulty of determining objectively what is an "attempt." Data on the incidence of suicide among psychiatric patients are sparse; they may be classified in two ways, according to occurrence: (*a*) after termination of treatment and (*b*) during hospitalization.

Posttreatment suicide rates among former psychiatric patients. Only four published studies could be located which reported data on the incidence of suicide among patients after termination of treatment. The authors of these reports obtained their data primarily by correspondence. The first three deal with psychoneurotics followed up for periods of from 2 to 20 years. Coon and Raymond [2] found 6 suicides among 1060 cases, a rate of .0057. Ross [17] reported 4 suicides among approximately 1186 neurotics, a rate of .0034. Denker [3] discovered 3 suicides among 707 cases, a rate of .0042. Denker's cases had not been psychiatric patients, but had been diagnosed as neurotic. A fourth study, by Holt and

---

[1] From the Neuropsychiatry Service, VA Hospital, Minneapolis, Minnesota, and the Divisions of Psychiatry and Clinical Psychology of the University of Minnesota.

Holt [10], reported 2 suicides among a mixed diagnostic group of 141 patients (a rate of .0142) within a period of 30 years after hospitalization. A third patient in this group committed suicide while under treatment subsequent to his initial hospitalization.

Suicide rates for hospitalized psychiatric patients. The rate of suicide among patients undergoing psychiatric treatment is apparently lower, probably because these patients are under close observation. Ross [17, p. 98] reported 3 suicides among 1186 neurotic patients, a rate of .0025. Two of these patients committed suicide during the period of treatment, and the other, immediately after termination. Levy and Southcombe [13] found 21 suicides among 6509 admissions (a rate of .0032) in a large state hospital for the period from 1936 to 1949.[2]

For the sake of the present discussion, a rough estimate of .0033 will be used to represent the suicide rate among psychiatric patients undergoing treatment. Although the rate may be slightly higher in some hospitals and clinics, it is evident that suicide occurs very infrequently among psychiatric patients.

Suggested Classification of Suicidal Patients

The need for more precise classification in research on suicide detection is thus emphasized both by the data on the low incidence of suicide and by the evidence earlier cited that "suicidal" subgroups can be differentiated by means of psychological tests. In the development and validation of a suicide detection device, one criterion group should consist exclusively of patients who committed suicide. Moreover, since a patient may undergo marked personality changes in the interval between test administration and the act of suicide, it is probably advisable to use data obtained only a relatively short time before the suicide. The most desirable time interval cannot be determined until more information on the stability of test responses of suicidal patients is available. The low incidence of suicide, of course, makes it extremely difficult to obtain presuicide data from an adequate number of cases for the development of a detection instrument.

Limitations of Suicide Predictors

A suicide detection instrument, to be effective, must identify a fairly large proportion of suicidal patients (true positives) and should not misclassify a large number of nonsuicidal patients (false positives). The low incidence of suicide is in itself a major limitation in the development of an effective suicide predictor, for in any attempt at prediction of infrequent behavior, a large number of false positives are obtained. To illustrate this point, a hypothetical suicide detection index will be devised.[3]

Development of a Hypothetical Suicide Detection Index

Assume that a suicide detection index was developed from test data obtained from a psychiatric patient population divided into two groups: patients who committed suicide during treatment (Suicide population), and patients who did not commit suicide during treatment (Nonsuicide population). Suppose that the steps in the development and validation of the index were as follows: (*a*) A random sample of patients was selected from each of these populations; (*b*) an index was devised which consisted of test data which significantly differentiated the two criterion samples; (*c*) all of the cases in the two samples were scored on the index; (*d*) a cutting line was established so that an equal percentage, say 80%, of the patients in each sample were correctly classified; (*e*) for the purpose of cross-validating this cutting line with new Suicide and Nonsuicide samples, every psychiatric patient over a period of years was scored on the index, and these scores were not divulged so that they could have no influence on the treatment of the patients.

Cross-validation of the index. On the basis of the estimated suicide rate of .0033, there would be approximately 40 suicides among 12,000 patients. Assume that, in cross-validating the predetermined cutting line, only a slight amount of shrinkage occurs so that 75% of the patients in each of the two populations can be correctly identified.[4] Table 1,

---

[2] Dublin and Bunzel [4, p. 311] reported a suicide rate of 42.3 per 100,000 psychiatric patients in New York State hospitals for the years 1919-1929. As the source for this rate, they cited an article which could not be located because the reference is incorrect.

[3] The writer is indebted to Dr. Paul E. Meehl for indicating the importance of an inverse probability approach to the problem of suicide detection.

[4] The figure of 75% is used because it probably represents the maximum effectiveness generally achieved with cross-validated prediction and classification devices. For example, Ellis and Conrad [5, p. 406], in evaluating the effectiveness of the personality inventories used for screening "maladjusted"

Table 1

Number of Patients in the Suicide and Nonsuicide Populations Identified by a Hypothetical Suicide Detection Index with Three Different Cutting Lines

Behavior predicted by index	Actual behavior				Total with predicted behavior
	Did commit suicide		Did not commit suicide		
	Number	%	Number	%	
A. 75% True Positives; 75% True Negatives					
Will commit suicide	30	(75%)	2990	(25%)	3020
Will not commit suicide	10	(25%)	8970	(75%)	8980
Total with actual behavior	40	(100%)	11960	(100%)	12000
B. 60% True Positives; 90% True Negatives					
Will commit suicide	24	(60%)	1196	(10%)	1220
Will not commit suicide	16	(40%)	10764	(90%)	10780
Total with actual behavior	40	(100%)	11960	(100%)	12000
C. 2.5% True Positives; 99.5% True Negatives					
Will commit suicide	1	(2.5%)	60	(.5%)	61
Will not commit suicide	39	(97.5%)	11900	(99.5%)	11939
Total with actual behavior	40	(100.0%)	11960	(100.0%)	12000

Part A, indicates the number of patients in each population who are identified by the index. In the Suicide population, only those patients whose scores are above the stipulated cutting line are correctly identified (true positives), whereas in the Nonsuicide population, only those whose scores are below this line are accurately classified (true negatives). The effectiveness of the index must be evaluated in terms of the number of correct identifications, in both populations, among all the patients predicted to be suicides.

It can be seen that with the use of the index 30 of the 40 suicides could have been correctly predicted. However, of the 11,960 patients in the Nonsuicide population, 2990 would also have been predicted suicides (false positives). Since every patient scoring above the cutting line would have been predicted to commit suicide, 2990 out of 3020, or 99% of such patients are misclassified. Obviously, such a suicide detection index would have no appreciable value, for it would be impractical to treat as suicidal the prodigious number of misclassified cases.

Type of prediction made with the index. In planning the initial development or cross-validation of any psychological instrument, it is essential to consider the kind of decision which is to be made with the test results. In the prediction of suicide, the test results must be evaluated solely in terms of the number of correct classifications among those patients predicted to be suicides. In Table 1, Part A, there were only 1% of such correct classifications. Among all the patients in the Suicide and Nonsuicide populations, however, 9000 out of 12,000, or 75% were correctly classified. It is not appropriate or meaningful to evaluate the index in terms of the latter percentage, for there is no need to identify the patients in the Nonsuicide population. In fact, without any test or other information, all patients could be predicted to be nonsuicidal, and the prediction would be correct in 99.67% of all cases.

Cross-validation of an elevated cutting line. For a suicide detection index to have any conceivable value, the number of false positives must be drastically reduced. This could have been effected in the original development of the index (step *d*) if a much higher cutting-line had been established. Assume, for example, that when this elevated cutting line was cross-validated (step *e*), only 10% of the nonsuicide cases scored above the line (i.e., were false positives). Assume also that the new cutting line reduced to 60% (still a liberal estimate) the number of correctly classified suicide pa-

---

from "normal" servicemen during World War II, found that, in general, roughly 75% of each group were identified.

tients. Table 1, Part B, illustrates the results of applying the index to the Suicide and Nonsuicide populations. Of the 40 patients who did commit suicide, 24 could have been identified by the index. There would still be 1196 false positives, however, so that if every patient who scored above the cutting line were called suicidal, the number of correct classifications would be only 24 out of 1220 or 2%. The index, therefore, would still be an impractical instrument because of the large number of false positives.

Further reduction of the number of false positives. With every elevation of the cutting line for the purpose of reducing the number of false positives, there will be a decrease in the number of correct predictions for the Suicide population (true positives). If the original cutting line were raised so that in the cross-validation procedure almost all false positives were eliminated, it might still be possible to detect one or two cases in the Suicide population by means of the index. Table 1, Part C, illustrates such a hypothetical situation. Although only 1 in 61, or less than 2% of the predicted suicides, is correctly identified, the important consideration is that the *number* of false positives has been reduced to a relatively manageable figure. The writer can think of no other event of relevance to clinical psychologists for which one might settle for such a low level of psychometric prediction. The attitude of hospital administrators, however, is that suicides must be prevented at almost any cost. Therefore, if the items comprising the index could be administered to all patients without undue expenditure of clinical time, the results in this apparently unique type of problem might be considered of some use.

Prediction within a Population Restricted to Certain Diagnostic Groups

It can be seen that it is very difficult to predict effectively any event with an extremely low *rate* of occurrence because of the large number of false positives. Another approach to the problem of reducing the number of false positives might be the development of an index within a population consisting only of those diagnostic groups with the highest rates of suicide. The advantage of this procedure, of course, is that the rate of suicide within this restricted population would be higher than that in the entire psychiatric patient population. Two immediate obstacles to the use of this approach are: (*a*) There is virtually no published information on suicide rates for given diagnostic categories. The writer could locate only one report [4, p. 311] which provided adequate data of this kind. (*b*) As a smaller *number* of suicides occur within a restricted population than in an entire psychiatric patient population, many years would have to elapse before a sufficient number of cases with presuicide data could be collected. It would be necessary, therefore, to conduct a coordinated research program in many installations.

Additional limitations of the restricted population approach. It would eventually be possible to obtain information on suicide rates within diagnostic groups, and it is also conceivable that a sufficient number of cases with presuicide data could be collected. Even if these two difficulties were surmounted, however, the procedure of developing and applying an index within a population restricted to one or several diagnostic groups has several limitations: (*a*) The rate of suicide would still be low; probably no group could be found with a suicide rate higher than .02, and there would still be an excessive number of false positives. (*b*) A considerable proportion of suicidal patients would not be included in the restricted population because of (*i*) the limited reliability of diagnosis, and (*ii*) the occurrence of suicide in diagnostic categories not included in the restricted population [4, p. 311; 13). (*c*) Variables differentiating suicide and nonsuicide patients would be more difficult to obtain within a somewhat homogeneous subgroup than within the whole psychiatric population.

Prediction Based on Clinical Judgment

The same kinds of difficulties which occur in psychometric prediction of suicide are encountered in prediction based on "clinical judgment." In clinical practice a large number of false positives are also obtained. The clinician, in referring to a patient as "suicidal," is willing to err on the safe side and treat a great many patients as though they were very likely to commit suicide.

Necessary Basic Research

It has been shown that the effectiveness of a suicide detection instrument is a function of (*a*) the *rate* of suicide within a given patient population (complete or restricted), and (*b*) the *percentage* of correct classifications both in the suicide and in the nonsuicide subgroups of the given population. Other things being equal, a given index will be most effective in the population with the highest suicide rate. Also, within a population with a given suicide rate, that index will be most effective which correctly classifies the highest percentage of cases in each of the population subgroups.

An essential prerequisite to the development of an effective detection device is the acquisition of a large body of information concerning the distinguishing characteristics of patients who commit suicide. To obtain such information, patients must be studied intensively by means of comprehensive social histories, behavior ratings, and psychological tests. The differentiating data thus collected may be utilized in either or both of the following ways: (*a*) to exclude from consideration as potential suicides a large proportion of psychiatric patients in order to obtain a restricted population with a higher suicide rate; (*b*) to find a large number of discriminating predictor variables which will correctly classify a high percentage of each of the suicide and nonsuicide subgroups.

Limitations in the Prediction of Other Infrequent Events

Clinical psychologists deal with virtually no other events as infrequent as suicide. One exception might be the prediction of homicide, for the incidence of homicide in the general population is about half that of suicide [20, p. 75]. The limitations of prediction discussed in this paper, however, apply in lesser degree to any other behavior as its rate approaches .50 in a population dichotomized according to the presence or absence of the behavior.[5]

The problem of suicide detection provides an effective illustration of the importance of considering the rate of occurrence within a given population of *any* event which is to be predicted. Ellis and Conrad [5] emphasized this point in their critical review of personality inventories used for screening purposes in the military services. They indicated that a large number of false positives were unavoidable because of the relatively small proportion of "maladjusted" individuals. They listed some studies which were misleading because the false positives were reported only in terms of percentages rather than the actual number of cases. Moreover, they reported some investigations in which the normal and abnormal samples were equal in number, and which were evaluated as though the proportions in the population were also equal, i.e., .50 rather than roughly .95 and .05. Ellis and Conrad indicated that such studies provide a spurious underestimate of the number of false positives in the population. Hunt [11, p. 214], in reviewing the Ellis and Conrad article, further elaborated on these points because they are generally overlooked.

In the clinical psychology literature, population rates of the events under study are rarely reported or considered. However, the effectiveness of any method for the prediction of the behavior of *individuals* cannot be evaluated properly without at least a rough estimate of the frequency of the behavior in the population being studied.

Summary

A number of investigators have proposed certain signs and configurations from psychological tests for the detection of suicidal patients. The purposes of this paper have been (*a*) to suggest greater refinement of the classification of patients in suicide research, (*b*) to emphasize some of the limitations inherent in the prediction of suicide, and (*c*) to indicate the general applicability of such limitations in the prediction of any other behavior or event of infrequent occurrence.

Classification of patients in suicide research. The need for more precise classification of patients in suicide research was emphasized both by (*a*) evidence that "suicidal" subgroups can be differentiated by means of psychological tests, and (*b*) data on the low incidence of suicide. Suicide is an extremely infrequent event, and very few of those patients who are severely depressed or who express suicide

---

[5] Elaboration of this point and other methodological problems will be included in a paper by Dr. Paul E. Meehl and the writer which is to be published.

thoughts or threats actually commit suicide. Thus, data from such patients cannot arbitrarily be grouped under the class term "suicidal." It was suggested that in the development and validation of a suicide detection device, one criterion group should consist exclusively of patients who committed suicide and for whom test data were obtained a relatively short time before the suicide.

Limitations of suicide predictors. The low incidence of suicide is in itself a major limitation in the development of an effective suicide detection device, for in the attempt to predict suicide or any other infrequent event, a large number of false positives are obtained (patients incorrectly classified as suicides). To illustrate this point, a hypothetical suicide detection index was developed and cross-validated within a psychiatric patient population. It was demonstrated that such an index would have no practical value, for it would be impossible to treat as potential suicides the prodigious number of false positives. With elevations of the cutting line of the index so as to reduce the number of false positives to a practical level, it was estimated that the number of correctly identified suicidal patients (true positives) would also be drastically reduced, perhaps to zero.

Another approach to the problem of reducing the number of false positives is the development of an index within a population consisting only of those diagnostic groups with the highest rates of suicide. The major limitations in the procedure of studying a population restricted to one or a few diagnostic groups are as follows: (*a*) The suicide rate would still be so low as to produce an excessive number of false positives. (*b*) A considerable proportion of suicidal patients might not be included within the population because of both the limited reliability of diagnosis and the occurrence of suicide in many other diagnostic groups. (*c*) It would be more difficult to obtain variables differentiating between suicidal and nonsuicidal patients within one or a few diagnostic groups than within the whole psychiatric population.

The difficulties in predicting events of low incidence are not restricted to psychometric prediction, but are likewise encountered in prediction based on "clinical judgment." A suicide detection device is not feasible until much more is learned about the differential characteristics of patients who commit suicide.

Limitations in prediction of other infrequent events. The problem of suicide detection provides an effective illustration of the importance of considering the frequency within a given population of any behavior or event which is to be predicted. The limitations of prediction discussed in this paper apply in lesser degree to any behavior as its rate approaches .50 in a population dichotomized according to the presence or absence of the behavior. The effectiveness of any method for the prediction of behavior for *individuals* cannot be evaluated properly without at least a rough estimate of the frequency of the behavior in the population being studied.

Received March 30, 1954.

References

1. Beck, S. J. *Rorschach's test*: Vol. II. *A variety of personality pictures.* New York: Grune & Stratton, 1949.
2. Coon, G. P., & Raymond, Alice F. *A review of the psychoneuroses at Stockbridge.* Stockbridge, Mass.: Austen Riggs Foundation, 1940.
3. Denker, P. G. The prognosis of insured neurotics. *N. Y. State J. Med.*, 1939, **39**, 238-247.
4. Dublin, L. I., & Benzel, Bessie. *To be or not to be: a study of suicide.* New York: Harrison Smith & Robert Haas, 1933.
5. Ellis, A., & Conrad, H. S. The validity of personality inventories in military practice. *Psychol. Bull.*, 1948, **45**, 385-426.
6. Farberow, N. L. Personality patterns of suicidal mental hospital patients. *Genet. Psychol. Monogr.*, 1950, **42**, 3-79.
7. Fisher, S. The value of the Rorschach for detecting suicidal trends. *J. proj. Tech.*, 1951, **15**, 250-254.
8. Hertz, Marguerite R. Suicidal configurations in Rorschach records. *Rorschach Res. Exch.*, 1948, **12**, 3-58.
9. Hertz, Marguerite R. Further study of "suicidal" configurations in Rorschach records. *Rorschach Res. Exch.*, 1949, **13**, 44-73.
10. Holt, W. L., Jr., & Holt, Winifred M. Long-term prognosis in mental illness: a thirty-year follow-up of 141 mental patients. *Amer. J. Psychiat.*, 1952, **108**, 735-739.
11. Hunt, H. F. Clinical methods: psychodiagnostics. *Annu. rev. Psychol.*, 1950, **1**, 207-220.
12. Kamman, G. R. The Rorschach method as a therapeutic agent. *Amer. J. Orthopsychiat.*, 1944, **14**, 21-27.

13. Levy, S., & Southcombe, R. H. Suicide in a state hospital for the mentally ill. *J. nerv. ment. Dis.*, 1953, **117**, 504-514.
14. Lindner, R. M. Content analysis in Rorschach work. *Rorschach Res. Exch.*, 1946, **10**, 121-129.
15. Rabin, A. L. Homicide and attempted suicides: a Rorschach study. *Amer. J. Orthopsychiat.*, 1946, **16**, 516-524.
16. Rosen, A., Hales, W. M., & Simon, W. Classification of "suicidal" patients. *J. consult. Psychol.*, 1954, **18**, 359-362.
17. Ross, Thomas A. *An enquiry into prognosis in the neuroses.* London: Cambridge Univer. Press, 1936.
18. Simon, W., & Hales, W. M. Note on a suicide key in the Minnesota Multiphasic Personality Inventory. *Amer. J. Psychiat.*, 1949, **106**, 222-223.
19. Ulett, G. A., Martin, D. W., & McBride, J. R. The Rorschach findings in a case of suicide. *Amer. J. Orthopsychiat.*, 1950, **20**, 817-827.
20. U. S. Department of Commerce, Bureau of the Census. *Statistical abstract of the United States.* Washington: U. S. Government Printing Office, 1953.
21. White, Mary A., & Schreiber, Hanna. Diagnosing "suicidal risks" on the Rorschach. *Psychiat. Quart. Suppl.*, 1952, **26**, 161-189.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

Intelligence Factors in Irregular Discharge Among Tuberculosis Patients[1, 2]

John R. Thurston and George Calden
VA Hospital, Madison, Wisconsin

Irregular discharge from tuberculosis hospitals is a problem of major medical and social importance. For each patient who remains until maximum hospital benefit has been achieved, there is another patient who leaves the hospital prematurely, without medical approval, i.e., receives an irregular discharge [1]. This suggests that the current tuberculosis therapy program can be applied effectively to only half of the hospitalized patients. From a public health point of view, the seriousness of this problem cannot be overemphasized.

It has often been hypothesized that low intelligence may be characteristic of the patient who leaves irregularly. This study was designed to answer the following question: Are patients who leave the hospital irregularly less intelligent than those who remain to receive a medically sanctioned discharge?

The *S*s in this study were 182 male tuberculosis patients admitted to this hospital during a one-year period. All *S*s were given a short-form Wechsler-Bellevue during their second month of hospitalization. The abbreviated form consists of six subtests: Information, Comprehension, Digit Span, Similarities, Block Design, and Picture Completion.

The irregular discharge group consisted of the first 50 *S*s in the sample to leave the hospital irregularly. The regular discharge group consisted of the first 50 *S*s in the sample who obtained regular, medically approved discharges.

Analysis of the W-B test results indicates that the tuberculosis patient group, as a whole, is of average intelligence (mean FIQ = 107.59).

A comparison of the mean IQ scores and subtest scores of the 50 regularly discharged patients (VIQ = 107.72, PIQ = 112.48, FIQ = 110.62) and 50 irregularly discharged patients (VIQ = 105.20, PIQ = 108.14, FIQ = 107.20) reveals no statistically significant differences. These results suggest that on a group basis, there is little difference in intelligence between patients receiving regular and irregular discharges.

In order to comprehend the relationship of intelligence to irregular discharge, it is probably necessary to consider the *interaction* of individual intelligence, personality, and situational factors, rather than to limit the evaluation to IQ scores per se.

Brief Report
Received July 19, 1954.

Reference

1. Tollen, W. B. Irregular discharge: the problem of hospitalization of the tuberculous. VA Pamphlet, 10-27, 1948.

---

[1] From the Veterans Administration Hospital, Madison, Wisconsin.

[2] An extended report of this study may be obtained without charge from John R. Thurston, Psychosomatic Section, VA Hospital, Madison, Wisconsin, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4340 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

Rorschach Scores, Prognosis, and Course of Illness in Pulmonary Tuberculosis[1]

David Cohen
VA Hospital, Coatesville, Pa.

In recent years, more and more attention has been given to the study of emotional factors as causative agents in and as reactions to the acute and chronic illnesses. Pulmonary tuberculosis, because of its chronicity and high incidence in the general population, has been the subject of many of these investigations with many further methodologically improved studies still in order. To date, various studies have been conducted and inferences drawn concerning possible personality factors predisposing to and maintaining the disease; the role of environmental stress in precipitating the disease; the psychological concomitants of the illness; relationship between personality traits and such variables as illness severity, course of the disease, the patient's acceptance of hospitalization, and the arrest and relapsing qualities of the disease; the selection of those patients most likely to leave the hospital against medical advice; and social and psychological factors in the rehabilitation of the tuberculous patient. Barker, Wright, and Gonick [1] make an excellent critical review of this literature. Other valuable critical reviews in this general area are offered by Bell, Trosman, and Ross [2, 3], and Windle [10].

Problem

The present study is an attempt to measure the ability of the Rorschach technique, singly, to predict two years beforehand the medical progress rate of the pulmonary tuberculosis disease process. Rate of medical progress is defined as the degree to which a patient's medical progress conforms to his expected progress based on the patient's over-all medical history. The medical progress ratings, as noted below, were independently made by three medical experts two years after the Rorschach protocols were obtained.

Procedure

The Rorschach protocols used in the present study were obtained from 45 male, white veterans hospitalized and being treated for active pulmonary tuberculosis at the Veterans Administration Hospital, Butler, Pennsylvania. The Rorschach data were originally obtained and analyzed by Newton [9] in his comparative study of the personality traits of patients with far advanced pulmonary tuberculosis as differentiated from those with minimally and moderately advanced tuberculosis. Twenty-four months after the original test administration, three physicians from the Tuberculosis Service of the hospital, upon the writer's request, rated the degree to which the course of the pulmonary tuberculosis process in each subject compared with the expected course of the disease in each case as determined by the entire medical history as shown in the medical chart. The rating scale instructions were specifically as follows:

> "Please rate on the attached scale the degree to which the course of the pulmonary tuberculosis process coincides with that which you would expect that disease process to have followed on the basis of the entire medical history as shown in the medical chart." The graphic rating scale consisted of five serially numbered points defined as follows: "Medical course much worse than expected"; "Medical course somewhat worse than expected"; "Medical course as expected"; Medical course somewhat better than expected"; "Medical course much better than expected."

In each case, a complete medical chart, as well as all X-ray plates, were available for analysis. All ratings by the three physicians were effected in a two-week interval so that an approximately equal period of time elapsed between the original Rorschach testing and the medical rating procedures. The three raters were highly competent medical practitioners specializing for many years in the diagnosis and treatment of tuberculosis. The three raters met together, reviewed the medical charts, X-ray plates, and other pertinent medical data, and each one then independently made his judgment on the above rating scale. Only those cases in which there was complete agreement between the three judges were included in this study. In each case, all ratings were made solely on the basis of the same data. None of the judges had close ward contact or immediate medical responsibility for the particular subjects of the study. This control was utilized in order to minimize personality bias on the part of the medical raters and to reduce the likelihood of their confounding adjustment to hospital routine with progress of the disease. The operation of this "halo" phenomenon and cautioning against it were pointed out by Daniels and Davidoff [4].

---

[1] From the Veterans Administration Hospital, Coatesville, Pennsylvania. This study was effected at the Veterans Administration Hospital, Butler, Pennsylvania.

Analysis of the Rorschach protocols was done independently of the medical ratings so that there was no contamination between medical ratings and test interpretation.

The *S*s of the present study had originally been selected in Newton's study in 24 pairs—each pair matched for age, number of months hospitalization for the disease, number of years education, Vocabulary score of the Wechsler-Bellevue Intelligence Scale, marital status, vocation, severity of the disease process, and whether the patient was improving, declining, or showing no significant change in his physical condition. No *S*s who showed any psychiatric pathology at the time of the study or in whose history there was evidence of marked emotional maladjustment were included.

One member of each pair was diagnosed as either a minimally or moderately advanced case of active pulmonary tuberculosis while the other member of the pair was diagnosed as a far advanced case of active pulmonary tuberculosis. The criteria of the National Tuberculosis Association [8] were routinely used by the medical staff in making this classification.

Table 1

Means, Standard Deviations, and Variance Ratios between Progress Groups on Age, Months of Hospitalization, and Vocabulary Score of Wechsler-Bellevue Intelligence Scale

Variable		Progress worse than expected	Progress as expected	Progress better than expected	*F*
N		19	15	11	
Age	Mean	29.84	26.87	25.64	2.42*
	SD	7.34	6.41	3.40	
Months of hospitalization	Mean	15.21	11.97	11.50	.88*
	SD	9.45	6.66	8.40	
Wechsler-Bellevue Vocabulary score	Mean	22.68	23.53	23.64	1.64*
	SD	5.18	5.17	4.25	

* Not significant at the .05 level of confidence.

Results

Table 1 presents the means, standard deviations, and variance ratios between the progress rating groups on age, Wechsler-Bellevue Vocabulary scores, and months of hospitalization. Degree of medical progress is not significantly related to any of these variables.

Since each *S* had originally been classified as to the severity of the tuberculous process, a contingency chi-square analysis was effected to determine the relationship between progress ratings and severity of illness. In a sense, such an analysis is a check upon the validity of the progress ratings since these ratings should not vary significantly with the severity of illness, in that the instructions to the raters were specifically to make their ratings in terms of everything known about each *S* including the variable of illness severity.

A chi square of 9.52, significant at $<.01$ $>.001$ level of confidence was obtained, indicating that the three raters were being significantly influenced in their progress ratings by the variable of illness severity. Further chi-square analysis between each of the progress rating groups indicated that the medical raters were assigning to the far advanced tuberculosis group a significantly greater number of "better than expected" progress ratings than "as expected" progress ratings. However, by combining the two "better than expected" groups with the "as expected" group, and comparing this combined group with the "worse than expected" progress groups, the variable of illness severity was found to be insignificant in the judges' progress ratings. With this analysis, the obtained chi square between progress ratings and the severity of illness variable was found to be 1.08, significant only at the .30 level of confidence.

These two derived progress groups were compared in terms of the Harrower-Erickson technique of evaluating neurotic signs in the Rorschach test as well as in terms of 35 Rorschach scoring variables.

Harrower-Erickson [6], by isolating and counting nine neurotic signs in Rorschach protocols, was able to identify 80% of the records of neurotic patients from the control subjects used in her study. Newton [9], using these signs, demonstrated a greater number of neurotic personality traits in the far advanced pulmonary tuberculous group than in the less advanced tuberculous group.

In the present study, the group rated as doing "as well as or better than expected" achieved a mean number of neurotic signs of 3.19; standard deviation 1.90. The group rated as doing "worse than expected" achieved a mean number of 3.68 of such neurotic signs; standard deviation 2.24. Analysis of variance reveals no significant difference between the groups in terms of the Harrower-Erickson neurotic signs.

The 35 Rorschach scoring variables studied were: W, D, d, Dd, S, R, M, FM, m, k, K, FK, $F\%$, $F+\%$, Fc, c, C', FC, CF, C, P, $R+\%$, percentage of responses occurring on the last three cards, average reaction time for chromatic, achromatic, and all ten cards, $H\%$, $Hd\%$, $A\%$, $Ad\%$, $At\%$, number of content categories other than the preceding four, number of card rejections, and the Anxiety and Hostility scores of Elizur's method of analyzing Rorschach content [5]. The last two scoring variables are based on Elizur's Rorschach Content Test procedure and have been shown to be related to the subject's covert hostility and covert anxiety, which are manifested in perceptual and fantasy living rather than being acted out.

The H test of Kruskal and Wallis [7] was used to measure the significance of differences between the two progress groups. The Kruskal and Wallis method is a ranking technique which assumes only that all observations are independent; that all observations within a given sample are derived from a single population; and that the sample populations are of approximately the same form. It makes no assumptions about scale units and the normality of score distributions, neither of which are met by the Rorschach data and are required if conventional additive statistical techniques were to be used. A derived H value is treated as a chi-square value with the degrees of freedom equal to the number of samples being compared less one. There is a correction for ties.

On none of the Rorschach scoring variables did the progress groups differ significantly. Apparently with patients of relatively young age, as in the population of this study, who have been hospitalized for active pulmonary tuberculosis for a relatively short period early in their illness, the Rorschach technique cannot reliably predict the rate of recovery at least in a twenty-four month interval.

Summary

This study measured the ability of the Rorschach technique, singly, to predict two years beforehand the medical progress rate of young, male, hospitalized, pulmonary tuberculous patients being treated early in their illness. Rate of medical progress was defined as the degree to which a patient's medical progress conforms to his expected progress based on the patient's over-all medical history. None of 33 Rorschach scoring variables, nor the Harrower-Erickson technique of evaluating neurotic signs on this test, nor the anxiety and hostility variables of Elizur's Rorschach Content Test were found to reliably predict the rate of medical progress. Degree of medical progress was not significantly related to age, length of hospitalization, intelligence, or severity of the tuberculous process.

Received May 3, 1954.

References

1. Barker, R. G., Wright, B. A., & Gonick, M. R. Adjustment to physical handicap and illness: a survey of the social psychology of physique and disability. *Soc. Sci. Res. Coun. Bull.*, 1946, No. 55.
2. Bell, A., Trosman, H., & Ross, D. The use of projective techniques in the investigation of emotional aspects of general medical disorders:

Part I. The Rorschach method. *J. proj. Tech.*, 1952, **16**, 428-443.

3. Bell, A., Trosman, H., & Ross, D. The use of projective techniques in the investigation of emotional aspects of general medical disorders: Part II. Other projective techniques and suggestions for experimental design. *J. proj. Tech.*, 1953, **17**, 51-60.
4. Daniels, G. E., & Davidoff, E. The mental aspects of tuberculosis. *Amer. rev. Tuberc.*, 1950, **62**, 532-538.
5. Elizur, A. Content analysis of the Rorschach with regard to anxiety and hostility. *Rorschach Res. Exch.*, 1949, **13**, 247-284.
6. Harrower-Erickson, Molly R. The value and limitations of the so-called "neurotic signs." *Rorschach Res. Exch.*, 1942, **6**, 109-114.
7. Kruskal, W. H., & Wallis, W. A. Use of rank in one-criterion variance analysis. *J. Ame statist. Ass.*, 1952, **47**, 583-621.
8. National Tuberculosis Association. *Diagnosti standards and classification of tuberculosi* New York: National Tuberculosis Associatio 1940.
9. Newton, R. A comparison of the Rorschac test records of far advanced tuberculosis pa tients with those of minimally and moderatel advanced tuberculosis patients. Unpublishe master's thesis, Univer. of Pittsburgh, 1949.
10. Windle, C. Psychological tests in psychopatho logical prognosis. *Psychol. Bull.*, 1952, **49**, 451-482.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

Test Indicators for Therapy Prognosis

James J. Gallagher
Michigan State College[1]

The recent extravagant use of psychotherapy as a cure-all in the area of emotional problems has rivaled the indiscriminate use of penicillin in its early days. One can push the analogy a step farther: while both have their legitimate areas of application, their usefulness has been clouded by the initial overselling of their potentialities.

The problem of psychotherapy today is not does it work, but rather, in what particular instances does it work. All of the various schools of psychotherapy have made lists of characteristics which distinguish a good therapy prospect from a poor one [1, 7, 11]. These lists, for the most part, have been developed from long clinical experience, but there is little objective evidence available.

Problem. An attempt was made to differentiate the most successful from the least successful therapy cases on the basis of pretherapy test indicators. The writer [9] has previously made a comparison of clients who drop out of therapy contact after one or two interviews with a group that continued contact. He found, as predicted, that overt stress seemed to be a determining factor in keeping clients in therapeutic contact. He also found that a lack of productivity on tests was characteristic of the group who discontinued therapy. The writer interpreted this lack of productivity as defensiveness on the part of the client representing an unwillingness or inability to fully enter into the total therapeutic situation.

The hypotheses in the present study are an extension of those tested in the previously mentioned research.

1. Measures of overt stress will differentiate the highly successful from the least successful therapy cases.
2. Measures of productivity will differentiate the highly successful from the least successful therapy cases.

Previous studies. One of the surprising findings in a perusal of the literature is the lack of prognostic studies of psychotherapy with standard psychological tools. The few studies that do appear have been more concerned with the prognosis of drug and shock therapy than with psychotherapy.

Bradway, Lion, and Corrigan [4] developed a prognostic rule on the Rorschach to predict the success in therapy of 16 promiscuous delinquent girls. When cross validated on 20 other delinquent girls, the list proved to be 80% successful in predicting the therapeutic outcome.

Harris and Christianson [12] attempted to match the pretherapy test results of 53 nonpsychotic patients to their success in brief psychotherapeutic contacts. The Wechsler-Bellevue scale and the traditional scoring of the Rorschach yielded no real differences between the highly improved group and the less improved group. The authors extracted Rorschach factors that seemed to be of significance in an attempt to form a prognostic scale. Unfortunately, no cross validation was done. Four MMPI scales, *Sc, Pa, Ma,* and *Pd,* revealed significant differences between the groups, with high scores being contraindications to a favorable outcome in therapy.

Barron [2] used the Rorschach, MMPI, and the Ethnocentrism scale to compare the results of 33 adult psychoneurotics following six months of psychotherapy in an outpatient clinic. Therapists' and supervisors' ratings were the criteria for improvement. The Rorschach revealed no differences between the improved and unimproved groups, but the MMPI profiles yielded results that agreed in part with the Harris and Christianson study. The unimproved groups scored higher on both the psychotic and neurotic scales on the pretherapy tests. The highest single predictor was the Ethnocentrism scale, a high score on the E scale being indicative of poor therapy prospects. Barron points out that the ethnic background of the therapists in this study might have had some untoward influence on this particular result.

---

[1] Now at the University of Illinois.

Rogers and Hammond [15] obtained a sample of 109 patients drawn randomly from a VA Mental Hygiene Clinic population. Success in therapy was rated by the therapist and senior psychiatrist on the staff. Neither clinical judgment on the Rorschach nor the Harris-Christianson scoring list showed any positive findings. Ninety-nine single variables were also examined with similar negative results. The only positive finding was the relatively high frequency of extensor *M* responses in the records of the successful group.

Method

Subjects. The 76 subjects involved in this study were all students of the Pennsylvania State College who came to the Psychological Clinic between September 1949 and August 1950 to obtain aid in their personal adjustment. A more detailed description of the subjects has already been given in previous papers [9, 10, 11].

Procedure. The 15 most successful clients and the 15 least successful clients were chosen on the basis of their scores on a multiple criteria scale for success in psychotherapy developed by Tucker [17]. This scale, which produces a single total score, consists of therapist ratings, judges' ratings, client ratings, and the ratio of positive to negative feelings in the first and last interviews.

Both the standard t test and the Mann-Whitney nonparametric technique [13] were used to test the significance of differences between groups. The significance of differences between the composite MMPI profiles of the Most Successful and Least Successful groups were determined by methods suggested by du Mas [5]. Since Barron [3] had developed an ego-strength scale which has shown considerable promise in predicting response to psychotherapy, the two groups were compared on this measure also.

Measures. The measures of overt stress were as follows:

1. The Taylor [16] anxiety scale. Since the short form of the MMPI was used, only 34 items were available for this study.
2. The Elizur [6] Rorschach anxiety scale.
3. The number of problems checked on the Mooney Problem Check List.

The measures of defensiveness or lack of productivity were:

1. The number of responses on the Rorschach.
2. The number of words used in summarizing their problem on the Mooney Problem Check List.
3. A defensiveness scale was developed on the MMPI in a previous study but proved to be too unreliable to be of any further use.

The results on the above scales for the entire therapy group have previously been presented by the author [9].

Table 1

Group Comparison on Measures of Stress and Productivity

Measure	Group					
	Discontinue ($N = 34$)		Least Successful ($N = 15$)		Most Successful ($N = 15$)	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Overt stress						
Taylor anxiety	8.81	5.39	14.50	6.04	18.53	5.52
Elizur anxiety	6.31	4.09	6.46	3.06	7.53	4.34
Mooney problems checked	33.86	25.65	42.08	35.15	47.00	28.29
Productivity						
Rorschach responses	20.74	11.51	21.93	9.66	26.60	11.38
Mooney words used*	46.54	12.23	46.83	6.73	53.14	10.87
Ego defense						
Barron scale			41.57	4.61	42.14	4.95

* Converted into T scores.

Results

Table 1 reveals the results obtained when comparing the Least Successful and the Most Successful therapy groups together with the results obtained previously on the Discontinue Therapy group, one in which the clients left therapy following one or two interviews. It may be noted that there is a constant progression from the Discontinue group to the Most Successful group on both overt stress and productivity. Nine comparisons made on measures of overt stress and six comparisons on verbal productivity show differences in the predicted direction in every instance. The correlations between the various scales were quite consistently of a low positive order (.25 to .35) and could not, in themselves, account for such consistency.

Table 2

t Tests of Differences between Means of Groups

Measure	Groups compared Discontinue vs. Least Successful	Discontinue vs. Most Successful	Least Successful vs. Most Successful
Overt stress			
Taylor anxiety	3.07**	5.59**	1.84*
Elizur anxiety	.14	.90	.75
Mooney problems checked	.79	1.50	.40
Productivity			
Rorschach responses	.36	1.62	1.17
Mooney words used	.10	1.84*	1.84*
Ego defense			
Barron scale			.32

* Significant at 5% confidence level.
** Significant at 1% confidence level.

Table 2 shows the *t* tests of differences between the means of the groups. The Taylor anxiety scale clearly is the most discriminating between the groups, and the number of words used in summarizing problems significantly separates the Most Successful group from the other two groups. Several of the other measures, e.g., the number of Rorschach reponses and the number of problems checked, approached but did not reach the commonly accepted levels of statistical significance.

The one measure that did not seem too effective was the Barron ego-defense scale, which yielded no differences between the Least Successful and Most Successful groups.

Table 3 shows the mean MMPI scores for the Least Successful and the Most Successful groups. Analyses of these results by du Mas' methods of profile analysis revealed the two profiles to be quite similar in shape, elevation, and scatter. The one difference that is statistically significant appears on the *D* scale on which the Most Successful group scored significantly higher before therapy began. The author has previously shown that this is one of the scales most likely to be influenced by contact with therapy.

Table 3

Comparison of Less Improved with More Improved on Pretherapy MMPI Scores

Scale	Less improved ($N=15$) Mean	*SD*	More improved ($N=15$) Mean	*SD*
F	60.86	8.66	59.67	7.99
K	54.27	7.56	51.73	6.63
Hs	56.33	12.54	56.27	13.20
D	65.73*	10.51	75.07	16.59
Hy	62.20	12.13	64.07	9.46
Pd	67.00	13.15	66.40	14.41
Pa	60.67	8.27	60.40	8.96
Pt	68.13	13.06	71.20	10.18
Sc	67.27	13.78	64.93	13.34
Ma	61.33	13.53	56.60	11.40

* Difference significant at 5% level of confidence.

The present MMPI results differ considerably from both the Harris and Christianson study and Barron's work. The major pretherapy differences on the profiles are the lower mean scores on the *Pa* and *Sc* scales in the present study. This result may be due to the weeding out of clients with psychosis or psychotic trends before starting therapy or may be due to the absence of those characteristics in the present college population.

Discussion

The results seem to justify a tentative conclusion that some of the factors that are related to clients leaving client-centered therapy are also operating to a lesser degree to prevent any substantial success in therapy itself. Two of these factors seem to be lack of overt tension and a lack of verbal productivity as revealed on the pretherapy tests. An interesting extension of this study would be to compare the relationship between lack of productivity in the tests and lack of productivity in the interview situation itself.

The implications here are that unless the client is motivated by some overt anxiety to change his perceptions and unless he is able to give of his perceptions freely with the counselor, then client-centered methods of counseling will produce a minimum of change whether or not the client leaves or remains in the situation. This, of course, does not necessarily mean that any other therapeutic methods would be more successful until evidence is presented to prove that contention.

It occurred to the writer after the data were collected that the statistical significance of the differences between the present groups on the stated variables depended on some assumptions that probably were not completely met. It is generally assumed that when groups are compared on one variable that the other relevant variables are either held constant or randomized. In the present case, that would mean that there would be as much liklihood for circumstances to cause clients to move from the discontinue category or the least successful category into the most successful category as there would be for them to go in the opposite direction.

Although no concrete evidence was available, it seems most unlikely that the above assumption was satisfied. A number of legitimate reasons can be given as to why clients who should have been in the Most Successful group according to our test predictors ended in the Discontinue group (moved out of town or failed out of school) or in the Least Successful group (poor counseling). It is hard to conceive of circumstances of equal probability that could turn discontinue cases or minimally successful cases into highly successful cases.

Thus, each of these adverse circumstances, present to an unknown degree, would have the effect of lowering the probability of getting statistically significant differences on the original variable. Therefore, any differences that are obtained should be viewed with a favorable eye, and the experimenter should hesitate before throwing out data that approach but do not reach statistical significance under these circumstances.

It is always well to try to estimate to what extent the present results can be generalized to other situations. A possible clue is the disagreement between the present results and the previous findings of Barron and of Harris and Christianson. As has already been noted, the present population, as revealed by MMPI results, is a rather homogeneous group.

One might conclude that the present results will hold only if other factors are kept relatively constant. When the total population becomes heterogeneous, such as might be encountered in an out-patient mental hygiene clinic, then other factors such as the amount of psychotic underlay, psychopathic trends, etc. may be more important in determining prognosis than the amount of overt stress and lack of productivity mentioned here.

The next question that needs to be answered is what can be done with the group, now partially identifiable, that does not seem able to profit very much from traditional methods of counseling. Here, perhaps, lies the most fruitful area of experimentation and research in the near future.

Summary

Results on pretherapy Rorschach, MMPI, and Mooney Problem Check List tests were used to compare 34 clients who discontinued, 15 clients who showed the least gain, and 15 clients who showed the most gain from client-centered counseling.

The findings seem to support the hypothesis that success in client-centered counseling seems to be positively related to the amount of overt stress and negatively related to the amount of pretherapy verbal productivity shown on the pretherapy tests.

Received March 17, 1954.

References

1. Alexander, F., & French, T. *Psychoanalytic therapy; principles and application.* New York: Ronald, 1945.
2. Barron, F. Some test correlates of response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 235-241.
3. Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327-333.

4. Bradway, Katherine, Lion, E. G., & Corrigan, Hazel. The use of the Rorschach in a psychiatric study of promiscuous girls. *Rorschach Res. Exch.*, 1946, **10**, 105-110.
5. du Mas, F. M. Quick methods for the analysis of the shape, elevation, and scatter of profiles. *J. clin. Psychol.*, 1953, **9**, 345-348.
6. Elizur, A. Content analysis of the Rorschach with regard to anxiety and hostility. *J. proj. Tech.*, 1949, **13**, 247-284.
7. Fenichel, O. *The psychoanalytic theory of neuroses.* New York: Norton, 1945.
8. Festinger, L. The significance of difference between means without reference to the frequency distribution function. *Psychometrika*, 1946, **11**, 97-105.
9. Gallagher, J. J. A comparison of clients who continue with clients who discontinue client centered therapy. In *Group report of a program of research in psychotherapy.* Pennsylvania State College: Psychotherapy Research Group, 1953. Pp. 21-38.
10. Gallagher, J. J. MMPI changes concomitant with client-centered therapy. *J. consult. Psychol.*, 1953, **17**, 334-338.
11. Gallagher, J. J. Manifest anxiety changes concomitant with client-centered therapy. *J. consult. Psychol.*, 1953, **17**, 443-446.
12. Harris, R. E., & Christianson, C. Prediction of response to brief psychotherapy. *J. Psychol.*, 1946, **21**, 269-284.
13. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50-60.
14. Rogers, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.
15. Rogers, L. S., & Hammond, K. R. Prediction of results of therapy by means of the Rorschach test. *J. consult. Psychol.*, 1953, **17**, 8-15.
16. Taylor, Janet A. The relationship of anxiety to conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81-92.
17. Tucker, J. E. Measuring client progress in client centered therapy. In *Group report of a program of research in psychotherapy.* Pennsylvania State College: Psychotherapy Research Group, 1953. Pp. 55-59.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

Implications of the Performance of Delinquent Girls on the Rosenzweig Picture-Frustration Study[1]

Julia R. Vane
Hempstead Public Schools, New York

A group of 95 delinquent girls and a group of 50 nondelinquent girls of similar age, intelligence, and socioeconomic background were given the Rosenzweig Picture-Frustration Study according to standard instructions.

The delinquents were from a state reformatory; mean CA 18.9 ± 1.6, mean IQ 91.5 ± 14.1. The nondelinquents were from the sixth term of a vocational high school; mean CA 16.9 ±.8, mean IQ 94.2 ± 6.2.

The results showed that the delinquents differed significantly from the nondelinquents and from Rosenzweig's norms in *E, I,* and *M,* and from the norms in *OD* and *NP*. The delinquent responses indicated less than average tendency to turn aggression outward, and a greater than average tendency to turn aggression upon themselves or to evade it. The responses also showed less concentration on frustrating objects and greater than average interest in solving the problem.

That the delinquents differed from the norms and nondelinquents was not surprising, but it was not expected that they would respond in a manner which would indicate that they were less aggressive than either the norms or nondelinquents.

Two other studies [1,3] indicate that when subjects wish to make a good impression on the P-F Study, the responses differ from the norms in the same manner as the responses of the delinquents. It would appear therefore that the delinquents were trying to make a favorable impression on this scale.

That this ability to make a good impression on the P-F Study is not dependent upon intelligence was revealed when the delinquents were divided into two subgroups and the results compared. One subgroup ($N = 45$), had a mean IQ of 79, the other ($N = 50$), had a mean IQ of 102. There were no significant differences between the groups in any of the categories.

The possibility that the response pattern shown by the delinquents may be a general delinquent pattern is suggested by the fact that the P-F responses of 110 women prisoners [2] deviated significantly from the norms in the same manner and almost to the same numerical degree as the delinquent responses.

On the basis of this and other studies mentioned [1, 2, 3] it seems inadvisable to use the P-F Study as a means of analyzing patterns of reactions of delinquents.

Brief Report
Received June 8, 1954.

References

1. Barrett, R. Some effects of the "job-testing" attitude on the Rosenzweig Picture-Frustration Study. Unpublished master's dissertation, Univer. of Pittsburgh, 1949.
2. Fry, F. V. A study of reactions to frustration in 236 college students and in 207 inmates of state prisons. *J. Psychol.,* 1949, **28,** 427-438.
3. Silverstein, A. B. Changes in performance on the Rosenzweig Picture-Frustration Study. Unpublished master's dissertation, New York Univer., 1951.

---

[1] An extended report of this study may be obtained without charge from Julia Vane, 80 Riverdale Rd., Valley Stream, N.Y., or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4345 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

Tactual-Kinesthetic Perception as a Technique for Diagnosing Brain Damage[1]

James W. Parker [2]
Brooke Army Hospital, San Antonio, Texas

It has been demonstrated that many of the current tests for brain damage are inadequate [1, 5, 11], although a few, particularly those stressing visual-motor performances, have proved effective. However, there are patients whose visual acuity is reported to be satisfactory, but who experience marked perceptual defects in other spheres as a result of brain injury [6]. Therefore, in view of DeJong's statement [4, p. 47] that the most difficult tasks for brain-injured patients are those requiring the subject to break away from old mental habits and adapt to unfamiliar situations, it would seem that techniques placing minimal emphasis upon visual perception should be of importance in evaluating these patients. In addition, the assertion by Strauss and Lehtinen [13] that different individuals stress different sense modalities in their everyday living would suggest that a patient, while actually suffering impairment, could manifest a "normal" performance on tasks requiring primarily the use of those modalities that he had emphasized throughout his life.

Simple tests of tactual-kinesthetic perceptual abilities have long been included among the diagnostic techniques of the neurologist; yet, in spite of a rather general recognition that brain-injured patients manifest impaired functioning in this sphere [4, 9, 14], carefully controlled research in the area has been meager. The few studies have been in most instances of limited application, either because of experimental design or the specific nature of the investigations. In some of these studies the patients have been inadequately diagnosed; in others, controls have been lacking or not comparable with the experimental subjects. In the main, these studies have been applied to special groups, such as young children or mentally deficient adults. Strauss and Lehtinen expressly state that ". . . although a comparative psychopathology of brain injury in children and adults shows many identities, some features . . . are manifestations of childhood" [13, p. 117] and that final conclusions regarding perceptual disturbances cannot be drawn because comparative studies with their tests, some of which accent tactual-kinesthetic performances, have not been made.

It was the purpose of the present study (*a*) to determine whether there is a significant difference in tactual-kinesthetic performances of brain-injured and non–brain-injured patients, (*b*) to investigate the possible relationship between tactual-kinesthetic task performance on the one hand, and the site and extent of the brain lesions on the other, and (*c*) to compare the patients' tactual-kinesthetic performances with their Bender-Gestalt test performances, emphasizing the relative discriminative power of the two types of measures.

Subjects

The experimental group consisted of thirty adult male hospitalized patients with definitely diagnosed brain injuries of fairly recent origin. Most of them had been injured within six to ten months of the time of this study. They ranged in chronological age from 21 to 48 years, with a mean age of 25 years. Their mental ages, obtained from the Shipley-Hartford Retreat Scale, ranged from 9 years, 6 months to 18 years, 7 months, with an average mental age of 14 years, 1 month. Patients with gross sensory and motor defects on the dominant side, neuropsychiatric patients, the noticeably mentally deteriorated, and patients with severe visual defects were not included in the investigation.

---

[1] Part of a dissertation presented to the Faculty of the Graduate School of the University of Texas in candidacy for the Degree of Doctor of Philosophy. The author wishes to express his thanks to Professors Philip Worchel, Wayne H. Holtzman, Hugh C. Blodgett, and Lloyd A. Jeffress of the Department of Psychology, University of Texas, and to Colonel Charles S. Gersoni, U. S. Army, for their helpful suggestions throughout the design and execution of this investigation.

[2] At Walter Reed Army Medical Center when this research was done.

The control group of 30 adult male patients who revealed no evidence of brain damage was drawn at random from general medical hospital wards. They ranged in chronological age from 19 to 40 years, with an average age of 23 years. Their average mental age was 15 years, 2 months and ranged from 10 years to 18 years, 7 months.[3]

Procedure

Each *S* was assured that the purpose of the examination was to facilitate understanding of his condition, and he was then administered the Bender Visual-Motor Gestalt Test, an examination frequently administered first in a battery because its innocuous appearance is believed to alleviate anxiety. Without knowledge of which were "organic" and which were "non-organic" records, the reproductions were scored according to the system devised by Pascal and Suttell [12]. To see if this test would differentiate brain-damaged from uninjured individuals, the results of the two groups were compared. In addition, Bender-Gestalt test performances were compared with those on tactual-kinesthetic tasks to determine whether the latter techniques would reveal the presence of brain damage in *S*s not detected by the Bender-Gestalt test.

Following the Bender-Gestalt test, and after testing each individual for stereognostic ability by having him identify several tactually perceived objects, *S*s were administered the tactual-kinesthetic tasks. The materials consisted of fourteen 8 × 11 inch wooden boards with various patterns on them made up of partially raised thumbtacks, half with smooth backgrounds, and half with backgrounds of thumbtacks embedded flush with the boards. In a pilot study four of the patterns had been matched with four other patterns; a "6", for example, was found to be of equal difficulty with a "9". In the present experiment one of each pair of these equated patterns had a background of thumbtacks, to determine whether the brain-injured group would experience greater difficulty than the controls in differentiating the figures from their backgrounds.

With the boards hidden from the *S*'s view, he was asked to explore their surfaces with his dominant hand until he was prepared to draw what he felt. It was stressed that the surface should be explored only as long as was necessary to be able to reproduce it accurately, and that while accuracy was of first importance, time was being recorded and would be considered. Furthermore, *S* was requested to describe the various surfaces while he was feeling them, and to make any comments that should enter his mind. Verbalizations were recorded, and notations were made concerning the manner of approach employed by *S*.

When the first procedure was completed for all the stimuli, the entire series was again presented to *S*, in the same manner except that he was not asked to reproduce the surfaces, but to feel them until he was ready to choose from a series of drawings the one most nearly like the surface he had explored. For each stimulus there were seven drawings from which to choose. These drawings were prepared in such a way that the brain-injured patient's alleged tendency to react concretely [8] would be elicited. The correct choice was always a straight-line drawing without the background represented; other choices consisted of drawings wherein an incorrect figure was made with dark spots representing tacks; and finally there were choices with incorrect straight-line figures on backgrounds representing tacks.

Methods similar to those of Strauss and Lehtinen with children [13] were used to compare the drawings made by the experimental group in the present study with those of the control group on the basis of the following criteria: (*a*) the number of straight-line drawings versus the number of drawings representing thumbtacks, (*b*) the number of drawings with only circles of the background, (*c*) the number of drawings with only the figures as well as the background, and (*d*) the number of figure rotations. In addition, two investigators independently rated each drawing on a scale from zero to five. A rating of five was given to drawings most accurately representing the stimulus figure and a rating of zero for the most inaccurate representations. In order that the judges would not know whether they were rating the record of a brain-injured or of a control *S*, all identifying data were removed and the records were shuffled during the rating procedure. The Spearman-Brown reliability coefficient for the two raters was .98. The ratings assigned to individual drawings by each judge were averaged separately, and the mean of the two averages was an *S*'s final total drawing score.

---

[3] The difference of one year, one month, in average mental ages between the two groups was not significant at the .05 level of confidence.

It was fully realized that the brain-injured *S*s of this study were not ideally suited for the investigation of the effects of cortical localization. The precise location and extent of such lesions are almost impossible to determine. Nevertheless, in view of the scarcity of previous studies in this area, it was decided to use the data to investigate effects of localization on tactual-kinesthetic performances. In all but four cases, wherein craniotomies were not performed, the brain surgeon or neurologist indicated on a chart, as well as he could ascertain, the brain areas involved in each lesion. The *S*s were divided into groups according to the site of their lesions, and analyses were made to determine whether certain brain areas were particularly related to tactual-kinesthetic performance. Finally, the data were examined for significant relationships between the extent of the lesions and the degree of disturbance noted.

Results and Discussion

Bender Visual-Motor Gestalt Test. The mean score of the brain-injured group on the Bender-Gestalt was 42.4, while that of the controls was 26.3. The difference of 16.1 between these means is statistically significant beyond the .01 level of confidence, thereby substantiating the assertion by Pascal and Suttell [12] that damage to the cortex results in higher scores on the Bender-Gestalt test.

When the median score of 28 was used as a cutting score, 72.4 per cent of the brain-injured group fell above this point, and 36.6 per cent of the uninjured scored higher. With a cutting score of 39, moreover, 48 per cent of the brain-injured *S*s were included in the higher group, and 10 per cent of the uninjured.

Not all of the very high-scoring patients were clinically apparent "organics." Some of them would undoubtedly have constituted diagnostic problems were it not for their history of traumatic brain injury. This observation does not corroborate the comment by Pascal and Suttell that only in extreme cases, which are also clinically apparent, can the Bender-Gestalt test answer the question: "Does the psychological record show signs indicative of damage to the cortex?" [12, p. 40].

Object-recognition test. Because the two groups could not be significantly differentiated on the basis of their ability to recognize nine different objects by the sense of touch, it would appear that tactual acuity in the brain-injured group was not seriously impaired, and that astereognosis was absent.

Analysis of the drawings. On the basis of the scores assigned to the various drawings of the stimuli, the experimental and control groups were differentiated beyond the .01 level of confidence. With the median of 35 as a cutting score, 76.6 per cent of the experimental group were included in the low-scoring half, and 26.6 per cent of the controls; 75 per cent of all *S*s would have been correctly diagnosed on this basis. A cutting score of 27, moreover, included 46.6 per cent of the brain-injured *S*s and 10 per cent of the uninjured.

Although the Bender-Gestalt test approached these results, with the median as a cutting score it detected only one brain-injured *S* not scoring below the median on the tactual-kinesthetic drawings. Low tactual-kinesthetic drawing scores, however, included three brain-injured *S*s not suggested by the Bender-Gestalt scores. The coefficient of correlation between the Bender-Gestalt and tactual-kinesthetic tests was .68.

Unlike the reactions of the brain-injured children studied by Strauss and Lehtinen [13], the adult brain-injured *S*s of the present study could not be differentiated from the control group on the basis of (*a*) the number of *S*s who drew straight-line forms; (*b*) the number who drew only the circles of the background; or (*c*) the number who drew the figure as well as the background. The reason these results are so different from those found with children may be accounted for by the assumption that perception in children differs from that in adults, and that, as Hebb postulates" . . . an early injury may prevent the development of . . . capacities that an equally extensive injury, at maturity, would not have destroyed" [10, p. 292].

When the drawings were analyzed for 90

and 180 degree rotations from the true orientation of the stimulus figures, it was observed that rotation occurred in the case of eight brain-injured *S*s and only one control *S*, the difference being significant at the .02 level of confidence. Were this trend to be borne out by further investigation, rotations on this task, as on the Bender-Gestalt test [3], may prove to be an important pathognomonic sign.

Reaction times. Although the brain-injured group scored lower than the control group on their reproductions of the tactual-kinesthetic stimuli, these injured *S*s had far more objective experience with the task. Whereas this group averaged 2.07 minutes per surface, the uninjured *S*s explored each surface on the average of only 1.35 minutes—a difference significant at the .02 level of confidence.

Allen and Krato [2] have speculated that the knowledge that he is being timed may exacerbate the brain-damaged patient's anxieties relating to his physical disability and serve to disrupt any train of organization, and that such disruption of an already impaired organizational capacity may well cause slower reaction times in the brain-injured. In this connection Goldstein maintains that brain damage "causes a rise of the threshold, and retardation of the excitation" [7, p. 599]. There occurs, he states, a reduction in receptivity, causing the patient to take much more time to react. Furthermore, the apparent need for several of the brain-injured *S*s to count each individual thumbtack even when advised that this was unnecessary may have been what Strauss and Lehtinen have described as ". . . an unconscious adaptation to order the external world in such a way that a positive result, and with it the satisfaction of achievement, can be obtained" [13, p. 27].

Multiple-choice recognition responses. For both groups of *S*s, choosing from a series of drawings the one most like the surface just explored proved to be a difficult task. However, the brain-injured group made markedly lower scores. Their mean number of correct choices was 7.3, while that of the controls was 10.4. The difference of 3.1 is significant beyond the .01 level of confidence. The nature of the drawings undoubtedly created particular difficulties for each group. The *S*s had to decide what was most essential in each choice: (*a*) the accurate representation of a figure made with a straight-line drawing, (*b*) an incorrect figure consisting of dots representing tacks, or (*c*) an incorrect figure having a background representing tacks. It was thought that the latter two types of responses were more concrete in nature and might, thereby, be made more frequently by brain-injured patients. Contrary to this expectation, however, only 68 per cent of the brain-injured patients' incorrect choices were of the "concrete" type, whereas 73 per cent of the incorrect choices of the uninjured patients were. This analysis, therefore, failed to reveal any particular concrete mode of approach that these brain-damaged patients may have been experiencing.

Comparison of thumbtack-background versus smooth-background stimuli. Four of the stimulus figures had been equated for difficulty with four others, and one set of these was given a background of randomly embedded flat thumbtacks while the other set was left with smooth backgrounds. The performances of *S*s on these two sets of stimuli were then analyzed in terms of (*a*) time spent exploring the surfaces, (*b*) scores made on the drawings of the stimuli, and (*c*) ability to make correct choices in the multiple-choice experiment. Not one of these variables showed a significant difference between the two groups. The stimuli with thumbtack backgrounds offered special difficulties to each group, and both groups performed faster and more accurately on the stimuli with smooth backgrounds. The brain-injured *S*s manifested practically the same enhancements on the smooth backgrounds, and the same degree of impairment on the thumbtack backgrounds as did the control *S*s.

Qualitative analysis of the tactual-kinesthetic performances. Although the present investigation offers no evidence that the brain-injured *S*s experienced greater difficulty than the uninjured in differentiating the figures from their backgrounds, and while no quantitative evidence suggesting an otherwise singularly concrete approach has been elicited, closer scrutiny of the various performances suggests that they may in fact have been more or less confined to a concrete type of behavior. In addition to the aforementioned tendency for some of the experimental *S*s to count all the tacks, apparently as if they could not perform at all without doing so, several of this group literally drew individual thumbtacks in profile in their attempted reproductions, seemingly being unable to arrange the separate aspects of a figure into an integrated whole. The immediate apprehension of this situation for these *S*s appeared to be simply thumbtacks in a board, and this is what they drew. Goldstein emphasizes that brain-damaged patients cannot transcend this "immediate apprehension of the given thing or situation in its particular uniqueness" [8, p. 3], and that their concrete approach causes them to be "directed by the immediate claims which one particular aspect of the object . . . makes" [8, p. 3]. Furthermore, he maintains that because this experienced uniqueness exerts demands upon the patient, it is practically impossible for him to realize other functions of the object, or to think of it as an example or representative of a larger whole.

Another feature of the brain-injured *S*s' performances was their marked tendency to remain silent while exploring the various surfaces, in spite of the repeated instruction to tell the examiner whatever came into their minds. It seemed that most of these *S*s were unable to perceive the stimuli and, at the same time, tell what they were doing or thinking. This effect, too, has been observed by Goldstein [8], who subsumes it under disturbances in abstracting. He asserts that the more difficult performances for the brain-injured are those which demand that they take account of what they are doing, and that such tasks will suffer more than those not requiring reflection.

Three dysfunctions listed by Goldstein would seem to account in part for many of the disturbances experienced by the brain-injured *S*s on the tactual-kinesthetic tasks of the present investigation: (*a*) an inability to "account for acts to oneself; to verbalize the account," (*b*) an inability to "hold in mind simultaneously various aspects," and (*c*) an inability to "grasp the essential of a given whole; to break up a given whole into parts, to isolate and to synthesize them" [8, p. 4].

Localization and extent of lesions. The drawing scores of the patients whose lesions could be macroscopically localized were categorized according to the brain areas involved. Because there are so few cases within each category, statistical analyses of the data were limited. The mean score for those whose lesions involved the parietal area was compared with the mean for those whose lesions involved frontal areas, and the difference between these means was not significant. In order to determine if there was a relationship between drawing scores and whether the lesion was in the dominant or nondominant cerebral hemisphere, the average score for patients whose lesions were in the dominant hemisphere was compared with that for patients who had lesions in the nondominant hemisphere, and the difference also was not significant.

Finally, to see if tactual-kinesthetic drawing scores were related to extent of brain damage, the records were divided according to whether the injury was large or small. Lesions which appeared to involve more than approximately one-fourth of one hemisphere were considered to be large, and those involving less were considered small. The average score for patients with large lesions was 22.3, and for those with small lesions, the average was 31.7. Although there was a difference of 9.4 in favor of the latter group, this was not significant even at the .20 level of confidence and, therefore, could easily have occurred by chance.

Summary

The purpose of this study was (*a*) to investigate the performances of brain-injured patients on tactual-kinesthetic tasks, primarily to discover whether such techniques could be used to differentiate brain-injured from non–brain-injured patients, (*b*) to determine if there is a relationship between tactual-kinesthetic task performance on the one hand and the site and extent of brain lesions on the other, and (*c*) to compare the patients' tactual-kinesthetic performances with their Bender-Gestalt test performances, with particular emphasis on the relative discriminative power of the two types of measures.

Sixty hospitalized adult male patients were examined, half having brain injuries of fairly recent origin and half revealing no evidence of neurological involvement. Following administration of the Bender-Gestalt, and after examining each individual for stereognostic abil-

ity, the *S*s were administered the tactual-kinesthetic tasks. The materials for these tasks consisted of 14 wooden boards with various patterns on them made up of partially raised thumbtacks, some with smooth backgrounds, and others with backgrounds of thumbtacks randomly embedded flush with the boards. With the boards hidden from the *S*'s view, he was asked to explore their surfaces with his dominant hand, to tell whatever came to mind, and then to draw a picture of what he had felt.

Although both the Bender-Gestalt and the tactual-kinesthetic techniques significantly differentiated the two groups, the latter technique appeared to offer greater discriminative power, differentiating the brain-injured from uninjured subjects at highly significant levels on the basis of the following criteria: (*a*) scores assigned to drawings of the stimuli, (*b*) the time consumed in exploring the various surfaces, (*c*) the responses to multiple-choice recognition items, and (*d*) the tendency to rotate reproductions of the stimulus figures. In view of these findings, it was concluded that tactual-kinesthetic perception, as has been shown with perception in other spheres, is often decidedly impaired in brain-injured individuals, and that performances relying on such perception may be used to differentiate brain-injured from uninjured patients.

No marked relationships were noted between tactual-kinesthetic task performances and the location or extent of the brain lesions.

Received April 13, 1954.

References

1. Aita, J. A., Armitage, S. G., Reitan, R. M., & Rabinowitz, A. The use of certain psychological tests in the evaluation of brain-injury. *J. gen. Psychol.*, 1947, **37**, 25-44.
2. Allen, R. M., & Krato, J. C. The test performance of the encephalopathic. *J. ment. Sci.*, 1949, **95**, 369-372.
3. Bender, Lauretta. A visual motor Gestalt test and its clinical use. *Res. Monogr. Amer. Orthopsychiat. Ass.*, 1938. No. 3.
4. DeJong, R. N. *The neurologic examination.* New York: Hoeber, 1950.
5. Eysenck, H. J., & Halstead, H. The memory function. I. A factoral study of fifteen clinical tests. *Amer. J. Psychiat.*, 1945, **102**, 174-179.
6. Gelb, A., & Goldstein, K. *Psychologische Analysen Hirn-pathologischer Fälle.* Leipzig: Barth, 1920.
7. Goldstein, K. The modifications of behavior consequent to cerebral lesions. *Psychiat. Quart.*, 1936, **10**, 586-610.
8. Goldstein, K., & Scheerer, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
9. Halstead, W. C. *Brain and intelligence: a quantitative study of the frontal lobes.* Chicago: Univer. of Chicago Press, 1947.
10. Hebb, D. O. *The organization of behavior.* New York: Wiley, 1949.
11. Manson, M. P., & Grayson, H. M. The Shipley-Hartford Retreat Scale as a measure of intellectual impairment for military prisoners. *J. appl. Psychol.*, 1947, **31**, 67-81.
12. Pascal, G. R., & Suttell, B. J. *The Bender-Gestalt test: quantification and validity for adults.* New York: Grune & Stratton, 1951.
13. Strauss, A. A., & Lehtinen, Laura C. *Psychopathology and education of the brain-injured child.* New York: Grune & Stratton, 1948.
14. Werner, H., & Strauss, A. A. Pathology of figure-ground relation in the child. *J. abnorm. soc. Psychol.*, 1941, **36**, 236-248.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

Productivity of Clinical Hypotheses on a Sentence Completion Test[1]

Rue L. Cromwell
The Ohio State University
and Richard M. Lundy
University of Wisconsin

Reviews of the literature on the sentence completion test [6, 9] indicate that the SCT is frequently used in the diagnosis of adult personality problems. Rotter and Willerman [7] used the test as a method of studying personality in the military setting; Machover [3] discusses its function as a part of the total diagnostic program in a psychiatric hospital; Stein [10] and Sacks [8] suggest methods of interpreting material produced on the SCT. Also, Sacks and Levy [9] have emphasized the use of the SCT as an instrument in forming clinical hypotheses about the emotional attitudes and mechanisms of the patient.

In these and other studies [e.g., 1, 2, 4, 5], sentence stems incorporated into tests have been selected on an a priori basis or from clinical experience, depending upon the purpose of the test. Though tests so designed may furnish hypotheses to the clinician about patients, the degree to which the various stems are used in making those hypotheses remains unknown. The impression of the authors is that some stems consistently produce many clinical hypotheses while others consistently do not. This research has been designed to find those stems which are consistently most productive. It is assumed that, with a given test length, the more useful the material is in making and supporting hypotheses, the better position the clinician is in when evaluating the patient.

Types of SCT instructions have been discussed by Rotter [6], Sacks and Levy [9], and Meltzoff [4]. Instructions emphasizing (*a*) "real feelings" and (*b*) speed are commonly recognized. These two kinds are studied here in relation to their effect on productivity of hypotheses for the test as a whole.

Procedure

Construction of test. Sixty-five stems were chosen for the test. This seemed to be the maximum number a hospital patient could complete in one session without undue effort and fatigue. Forty stems were taken from the Rotter Incomplete Sentences Blank, Adult Form, because a great deal of research has already been done on them in other situations. The remaining 25 stems were created to secure further information about ward adjustment of patients,[2] patient identification, and the utility of stems using the third person pronoun.

Two different sets of instructions were used. They were alternated from patient to patient. One set emphasized speed and the other emphasized "real feelings" as in the Rotter ISB, as follows:

> *Set S.* This is a test of your *speed* of thinking. Complete the following sentences as *quickly* as you can. Write down on each line the first idea that occurs to you. It makes no difference whether the statement you write is true or not. Do not start until the examiner tells you to begin.
>
> *Set F.* Complete these sentences to express ***your real feelings.*** Try to do every one. Be sure to make a complete sentence.

With the subjects who received the speed instructions a stop watch was used as a prop.

The order of the sentence stems, as well as the instructions, was varied from subject to subject. The stems were reproduced on five pages, 13 stems to a page. The order of the pages was systematically alternated throughout both subject samples. By this means the subjects, and the clinical judges also, did not consistently respond to the same stem order on every test. For the second sample of 30 subjects the stems were reshuffled and a new order was arranged on each page. This complete rearrangement of stem order was made to prevent context variables from affecting the results.

---

[1] Acknowledgments are offered to the staff and trainees of Chillicothe VA Hospital and to the VA psychologists throughout the country who assisted in this research.

This paper is published with the permission of the Chief Medical Director, Department of Medicine and Surgery, Veterans Administration, who assumes no responsibility for the opinions expressed or conclusions drawn by the authors.

[2] Another study is being planned to investigate the use of the SCT in predicting ward adjustment.

Subjects. The SCT's, thus constructed, were administered to 60 patients soon after their admission to Chillicothe VA Hospital. The subjects were divided into two samples of 30 each (Sample A and Sample B). No regard was given to the reason for admission or the diagnosis. If a patient was too disturbed or uncooperative to complete as many as 55 of the 65 stems, he was not used in the research.

Judges. After the protocols had been completed by the patients, they were sent in pairs together with a standard set of instructions to clinical psychologists who acted as judges. Thirty-five judges were selected by geographical stratified sampling of VA neuropsychiatric hospitals throughout the country. Four judges were selected locally. This method of selection afforded a wide sample of theoretical approaches currently used among VA psychologists.[3] Each of the 39 judges evaluated either one or two of the 60 protocols. No judge judged protocols from both Sample A and Sample B.

The judges were asked to make hypotheses about the subject from the completed protocol in the same way they would if they were writing a case report of the individual. They were then asked to list these formulated hypotheses and to put after each one the code numbers of the completed sentences used in deriving the hypothesis. Inferences about ward adjustment were also made and listed separately.

Results

From the lists of personality inferences returned by the judges the number of times each judge used each stem was tabulated. No attention was given to the nature of the hypotheses or the number of stems used for a single hypothesis. The number of times any judge referred to any one stem ranged from zero to six times.

The number of times an individual judge referred to completed stems in a single protocol ranged from 15 to 112. If a simple frequency count were used in evaluating the productivity of individual stems, the judges making frequent references to stems would have a greater influence on the results than those judges who made infrequent references to them. To control this factor, the number of times each stem was used by each judge was divided by the total number of times he made references to *all* stems. The resulting weighted scores allowed each judge of each protocol to contribute equally to the results.

The weighted scores were summated for each of the 65 sentence stems for both Sample A and Sample B, giving each stem two productivity scores. These two sets of scores were correlated to see if the stem productivity was consistent from one sample to the other. The product-moment correlation was +.745.

The next step was to find a cutting point in the data in order to select the most productive stems for a final test form suitable for administration. By inspection, a cutting point was determined between the 45th and 46th stem, arranged in order of productivity. Of the 45 best stems in Sample A, 41 were also among the best 45 of Sample B. A rank-difference correlation of +.401 was obtained between the ranking of the best 45 in Sample A and their ranking in Sample B. This correlation is considerably below the correlation for all 65 stems, thus indicating that some stems which were consistently poor in productivity were eliminated. This final selection of 45 stems represents a compromise between the goals of high productivity and optimal test length for behavior sampling.

In order to obtain the most reliable ranking, the productivity scores from Samples A and B were added together for each stem. With the Spearman-Brown formula, we can estimate the correlation between these final productivity scores (combined from A and B) and the productivity scores from a proposed second set of judges and protocols similarly obtained. This inferred correlation coefficient, computed as +.85, is the estimated reliability of this final ranking of the stems:

---

[3] Among these 39 judges, 7 called their approach psychoanalytic; 13, modified, broad, eclectic, or neo-psychoanalytic; 8, eclectic; 4, learning or social learning; 3, Sullivanian; 1, Jungian; 1, client-centered; 1, neo-Adlerian; 1, perceptual-cultural.

1. People who push me around
2. My greatest worry
3. My greatest fear
4. I like to break
5. Marriage
6. The only trouble
7. The future
8. Sometimes
9. My mind
10. I'll try to
11. I suffer
12. I like
13. A mother
14. I need
15. I am very
16. I failed
17. I can't
18. Other people
19. Most women
20. I secretly
21. My father
22. I regret
23. I
24. People
25. She
26. He
27. I am best when
28. I hate
29. I feel
30. What annoys me
31. Fighting
32. When I was a child
33. I should
34. The happiest time
35. He wants
36. I want to know
37. I wish
38. What pains me
39. I should obey
40. The best
41. What worried us was
42. My nerves
43. Men
44. My friends
45. At bedtime

46. Back home
47. Our
48. Telling the truth
49. It
50. This place
51. In the service
52. They will
53. In school
54. Dancing
55. She needs
56. In the past we
57. They
58. AWOL
59. My superior officer
60. They had
61. Negroes
62. It seemed very
63. Reading
64. Where I worked
65. Sports

The final step in the analysis was to compare the "real feeling" instructions with the speed instructions in terms of total test productivity. The difference was not significant. However, the design did not permit comparison of instructions from ratings by the same judge. This finding, therefore, may have resulted from uncontrolled judge variability that masked real differences between instructions.

Discussion

In general, the results have supported the hypothesis that there are consistent individual differences in the productivity of sentence completion stems. The choice of stems for a test does make a difference as to how valuable the test is in sampling behavior. The correlation of +.745 between stem productivity scores of two subject samples attests to the consistency with which some stems are more productive than others.

By setting the cutting point at 45 and dropping the poorest 20 stems we derived a test form that seems to be of appropriate size for personality evaluation in neuropsychiatric hospitals. It extracts a good deal of information from the average patient without becoming a burdensome task. Nevertheless, the test size could be shortened or lengthened to meet the needs of the individual clinician. For a short screening blank the first 15 stems could be used. Ten of the first 15 stems in Sample A were among the most productive 15 in Sample B. The test could also be used flexibly as a time-limit rather than a work-limit test. The patient could start at the beginning and work until it is convenient for him to stop.

In this empirical investigation there arises the question of what qualities make a stem productive or nonproductive. To pursue this question, the first ten stems may be compared with the last ten stems in the above list. It appears from examining these extremes of productive and nonproductive stems that no single factor differentiates the two groups. Many of the good stems refer to the first person (I, me, my), whereas many of the poor stems have third person or impersonal references. Also, the poor stems often refer to activities with little "emotional" involvement. The good stems, on the other hand, are often built around hostilities, worries, troubles, and fears. Finally, the poor stems refer more often to past situations, whereas, the productive stems deal more with the present and the future.

It is interesting to note that the SCT does not appear to be a fertile technique for investigating past and childhood experiences. Not only are past tense references prevalent among the poor stems listed above, but stems such as "Back home" and "When I was a child" are low in the list. Whether it is because of the patients' ability to respond, the interests of the clinician in using the test, or the inherent nature of the stems, it seems that the SCT in the hospital situation is more fertile in sampling statements of present attitudes and future goals.

A next step in the development of this SCT would be a validation of clinical hypotheses derived from it. Such a study would yield a final SCT that is both productive and of known validity.

Summary and Conclusions

An incomplete sentences test of 65 stems was given to 60 newly admitted VA neuropsychiatric hospital patients. The order of the stems and two different sets of instructions were alternated from subject to subject. Thirty-nine clinical psychologists in various VA hospitals made personality inferences from the protocols and listed the completed sentences used in making the inferences.

The major findings were:

1. A correlation showed the individual differences in stem productivity to be significantly consistent from the first sample of 30 subjects to the second sample of 30 subjects.

2. A test of 45 highly productive stems was devised as appropriate for psychological evaluation in neuropsychiatric hospitals.

3. Results suggested that many of the more productive stems referred to the first person, the present and future, and "emotional" aspects of the subject.

4. No difference was found between speed and "real feeling" instructions with respect to test productivity.

Received April 1, 1954.

References

1. Hadley, J. M., & Kennedy, V. E. A comparison between performance on a sentence completion test and academic success. *Educ. psychol. Measmt.*, 1949, **9**, 649-670.
2. Lehner, G. F. J. Projections of men and women to items referring to the same and the opposite sex on a sentence completion test. *Amer. Psychologist*, 1947, **2**, 407. (Abstract)
3. Machover, S. (Ed.) Case reports in clinical psychology. Brooklyn, N.Y.: Kings County Hospital, Vol. 1, No. 1, August 1949.
4. Meltzoff, J. The effect of mental set and item structure upon response to a projective test. *J. abnorm. soc. Psychol.*, 1951, **46**, 177-189.
5. Rosenberg, S. Some relationships between attitudes expressed toward the parent in a sentence completion test case history data. *J. proj. Tech.*, 1950, **14**, 188-193.
6. Rotter, J. B. Word association and sentence completion methods. In H. H. Anderson & Gladys L. Anderson (Eds.), *An introduction to projective techniques*. New York: Prentice-Hall, 1951. Pp. 279-311.
7. Rotter, J. B., & Willerman, B. The Incomplete Sentences Test as a method of studying personality. *J. consult. Psychol.*, 1947, **11**, 43-48.
8. Sacks, J. M. Effect upon projective responses of stimuli referring to the subject and to others. *J. consult. Psychol.*, 1949, **13**, 12-21.
9. Sacks, J. M., & Levy, S. The Sentence Completion Test. In L. E. Abt & L. Bellak (Eds.), *Projective psychology*. New York: Knopf, 1950. Pp. 357-402.
10. Stein, M. I. The use of a sentence completion test for the diagnosis of personality. *J. clin. Psychol.*, 1947, **3**, 47-56.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

The Validity of MMPI Interpretations[1]

Kenneth B. Little
Stanford University
and Edwin S. Shneidman
Neuropsychiatric Hospital, VA Center, Los Angeles

In clinical psychology, diagnostic tests are the most common source of personality descriptions. The descriptions range from short paragraphs, emphasizing a single salient characteristic of the subject, to elaborate many-paged analyses of several aspects of his psychological functioning. Very rarely are the conclusions drawn by the clinician restricted simply to a diagnosis. The protocol, whether in the form of the set of scores or a profile from a personality inventory or the verbatim responses to projective materials, leads the interpreter to make a large number of inferences, only one of which is a diagnosis qua diagnosis. These inferences seldom have a direct overt relationship to objective characteristics of the protocol; they are a function of the interpreter's experience, skill, personality, etc., as well as of the test itself.

This situation is taken somewhat for granted with projective tests but is frequently ignored in considering the more objective personality measures such as the MMPI. The pristine beauty of the quantitative scores on this inventory has generally led investigators into testing a single inference, i.e., diagnosis, when judging the test's validity. Yet in practice the MMPI is used quite differently. The clinician inspects the profile, occasionally scores additional scales, perhaps examines the actual responses to individual items, and somehow during the process arrives at a number of conclusions which he embodies in a formal or informal psychological report. The validity, in clinical practice, of the instrument would therefore seem best reflected by some measure of the number of correct inferences arrived at by the clinician. The present paper is the report of a study of the validity, as just defined, of the MMPI.

It is obvious that an index of validity as described above will vary with the clinician who does the interpreting and that no unique figure will be obtained. The authors have suggested in a previous paper [1] that the central tendency of such indices among competent clinicians interpreting a variety of subject protocols might be considered the validity of the instrument itself. In this study, the results for a number of interpreters working with a single protocol are presented. They indicate only what *can* be achieved and not necessarily what might occur with other protocols. In addition, the results were examined to determine the types of inferences made on the basis of the MMPI and the relationship between the resultant types and their accuracy.

Procedure

The general procedure has been described in more detail elsewhere [1] and is presented here in synoptic form.

Eleven psychologists, competent in the use and interpretation of the MMPI and accustomed to writing psychological reports from it, were presented with the MMPI profile of a subject identified to them as male, age 25, and single.[2] On the basis of the information conveyed to them by this profile, each made a Q sort [3] of 150 items of personality description. The continuum was from "Most True" to "Most False" for the subject. The items were a stratified sample from a total item population of 1604 items abstracted from 17 psychological reports written about the subject on the basis of his TAT and MAPS test protocols. The 1600 items covered,

---

[1] Presented with the approval of the Chief Medical Director of the Veterans Administration. The statements and conclusions published by the authors are a result of their own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

This investigation was supported by a research grant from the National Institute of Mental Health of the National Institutes of Health, Public Health Service, administered by the University of Southern California.

[2] The test materials and case history data of the subject are presented in *Thematic Test Analysis* [2].

so far as could be determined, the usual aspects of psychological functioning presented in psychological reports.

The intercorrelations of the distribution of items for the 11 MMPI judges are presented in Table 1.

Table 1

Correlations among the Q Sorts of 11 MMPI Interpreters*

Interpreter	2	3	4	5	6	7	8	9	10	11
1	63	62	50	65	73	56	68	64	56	67
2		68	56	62	66	51	68	72	61	55
3			62	56	50	55	64	68	51	49
4				48	45	52	60	66	47	49
5					68	42	72	51	67	62
6						54	63	61	61	67
7							50	64	43	48
8								60	66	56
9									52	55
10										58

* Decimal points omitted.

The criterion consisted of the consensus of 29 experienced clinical psychologists and psychiatrists who Q sorted the same 150 items, except that their sorting was made on the basis of a complete clinical folder about the subject. The folder contained medical examination reports, laboratory data, course of treatment notes, psychotherapy notes, social history, consultation reports, etc., but excluded the psychological test reports.

A factor analysis of the intercorrelations of the 29 criterion judges [1] indicated that a single general factor would account for 90% of the communality among them. Accordingly, a new Q sort was constructed for the criterion itself, as follows:

1. A multiple-regression equation for estimating the criterion general factor was derived using the criterion judges' general factor loadings as correlations. It was found that with only three variables (judges) a correlation of .925 with the general factor was secured.
2. The Beta coefficients for these three criterion judges were used to weight each of their Q-sort scores for the 150 items. The three weighted scores for each item were summed to form a new composite score for that item.
3. The 150 items were then ordered into the criterion Q sort on the basis of their rank position in the distribution of composite scores.

To determine the validity coefficients, the 11 MMPI judges were correlated with this composite criterion Q sort.

Results

A cluster analysis [4] of the matrix of intercorrelations of the MMPI interpreters (Table 1) yielded three groups of judges. Table 2 gives the estimated correlations of each judge with each of the three cluster domains and also their communalities. (The correlation of a judge with the domain of which he is a member is analogous to a factor loading in a factor analysis.) Table 3 gives the correlations among the cluster domains themselves.

Table 2

Estimated Correlations of Each MMPI Interpreter with the Cluster Domains and the Validity Coefficient for Each Interpreter

Cluster	Judge	Cluster domain A	B	C	h^2	Validity
A	1	*.84*	.77	.73	.71	.72
	6	*.84*	.77	.64	.71	.74
	11	*.81*	.71	.63	.66	.70
B	5	.78	*.84*	.64	.71	.71
	8	.75	*.83*	.76	.69	.73
	10	.70	*.81*	.62	.66	.58
C	3	.65	.69	*.80*	.64	.70
	4	.58	.62	*.79*	.62	.56
	9	.72	.66	*.83*	.69	.66
	2	.74	.77	*.81*	.66	.70
	7	.63	.54	*.71*	.50	.52

Note.—Italicized figures are the correlations of the judges with the domain of which they are a member.

Table 3

Estimated Correlations among the Cluster Domains

Cluster domain	A	B	C
A	1.00	.90	.80
B	.90	1.00	.81
C	.80	.81	1.00

To test the hypothesis that the three clusters were sufficient to account for all the common variance among the judges, a table of theoretical correlations was computed on the basis of each judge's correlation with the cluster of which he was a member and the correlations among the domains (Tables 2 and 3). A comparison with the 55 empirical correlations of Table 1 indicated that 44 (76%) of the theoretical correlations were within one standard deviation of the corresponding empirical ones, 54 (98%) were within two standard deviations, and only one theoretical correlation deviated from its empirical counterpart by more than two standard deviations. The mean absolute discrepancy between the two sets of correlations was .024. It appears, therefore, that the three clusters do adequately account for the common variance in the matrix.

To identify the groups of judges, an average *Q* sort was made for each of the clusters. The 20 Most True items in each average *Q* sort were examined, and those items that appeared for one and only one cluster were isolated. The same process was also repeated for the 20 Most False items of the three clusters.

The *Most True* items for cluster A were:

He is in the early stages of paranoid schizophrenia.
He has strong latent homosexual feelings.
The psychological threat to him is really of a homosexual nature.
He has an enormous amount of hostility.
He would be threatened by interpretations given early in psychotherapy.
His principal conflict is in the sexual area.
He has at least bright normal intelligence.

The *Most False* items for cluster A were:

He does not have much anxiety.
He should be able to progress satisfactorily in psychotherapy without excessive support.
Masturbation creates no psychological problems for him.
He seems little concerned with his bodily well-being.
His ideation is not paranoid.

The *Most True* items for cluster B were:

He appears to be a solitary person.
An acute break with reality has probably occurred.
The total picture is consistent with a schizophrenic disorder with potential paranoid and hebephrenic coloring.
He avoids social disapproval by engaging in solitary activity.
He has never learned to solve his problems by other than avoidant means.
He has never adequately learned social skills.

The *Most False* items for cluster B were:

He appears to be an outgoing person.
For him, withdrawal is a relatively unimportant mode of reacting to frustration.
He has little guilt feelings concerning his aggressive impulses.
He has the well-preserved mind of a person who is not psychotic.
He is probably not disoriented.

The *Most True* items for cluster C were:

He perceives the world as consistently unloving.
He has marked guilt feelings.
His guilt feelings may overwhelm him.
He feels deprived of the oral gratifications of childhood.
He suffers from feelings of rejection.
In fantasy he longs for kind parents.

The *Most False* items for cluster C were:

He feels generally that he is master of his own fate.
He has little orality.
His superego is relatively mature.

This technique of identification emphasizes differences among the clusters; the actual high degree of agreement is apparent from the correlations among the cluster domains (Table 3). However, inspection of the cluster-identifying items suggests that the judges of cluster A had a slight preference for emphasizing the nature of the conflicts (over sexual impulses) of the subject; the judges of cluster B had a preference for emphasizing the primary defenses (withdrawal) of the subject; and the judges of cluster C had a preference for emphasizing the affective reactions (feelings of rejection) of the subject. All are describing a person who is essentially schizophrenic.

The right-hand column of Table 2 gives the validity figure for each judge in the form of a correlation with the composite *Q* sort described previously. The mean value of this validity figure for cluster A was .72, for cluster B, .67, and for cluster C, .63, with an over-all value for the 11 judges of .67. An analysis of variance indicated that the differences among the three clusters were not significant.

Discussion

The data presented above indicated that substantial agreement may occur between descriptions of an individual based upon his MMPI profile and those based upon an elaborate clinical history. The degree of agreement can be emphasized by pointing out that the average general factor loading (or correlation with the criterion) of the criterion judges themselves was only .74 [1] as compared to the average of .67 for the test judges with the composite criterion

Q sort. Moreover, the former figure is a correlation with a criterion of unit reliability whereas the latter is with an estimate of that criterion only. Correction of the average validity figure of the MMPI judges for the unreliability of the "true" criterion estimate would further decrease the difference between the two figures cited.

Certain cautions need to be kept in mind in evaluating these data, however. The results indicate only what *can* be done by competent clinicians with a specific protocol and not what might occur with different interpreters or with the same interpreters and different protocols. The subject used in this study presented an ambiguous clinical picture at the time of testing but his MMPI profile appears to be anything but equivocal (see Table 4). A strong possibility exists, therefore, that the comparison of the MMPI *Q* sorts with the criterion *Q* sorts tested the agreement of two sets of statements about a certain nosological classification rather than inferences about a specific person.[3] One critic was so unkind as to suggest that the data merely demonstrate the existence of a common delusional system among test and criterion judges. However, insofar as the statements made have a descriptive utility, agreement between test interpretations and those made on the basis of exhaustive clinical material seems a desirable form of validity no matter how the statements are derived.

Table 4

MMPI Scores of the Subject

Scale	Raw score	Standard score
?	68	60
L	2	43
F	13	73
K	10	46
Hs	22	90
D	36	95
Hy	35	85
Pd	27	79
Mf	36	80
Pa	21	88
Pt	43	111
Sc	49	120
Ma	25	75

Three clusters of judges could be found among the 11 judges used in this study, but the differences among them do not seem to be related to their over-all validity. This would follow logically from the discussion set forth above. The variations in description represent different emphases upon certain aspects of the same disorder rather than disagreement as to diagnosis. Thus the average validity of the three clusters is about the same.

A final conclusion from the study is that the *Q* technique has considerable value in the study of the reliability and validity of clinical techniques. It permits the quantification of results without loss of the idiographic approach, a characteristic ideally suited to clinical research.

Summary

The validity of inferences made from the MMPI was tested using an interpreter population of 11 experts working with the MMPI profile of one subject. The criterion was the consensus, in the form of a general factor, of 29 clinicians who *Q* sorted 150 items on the basis of a comprehensive clinical record of the same subject. A cluster analysis of the matrix of intercorrelations among the MMPI interpreters yielded three groups of judges. These groups were described in terms of the Most True and Most False *Q*-sort items peculiar to each cluster. Validity figures for each MMPI judge were computed as correlations with a weighted composite *Q* sort of the criterion judges. The general results indicated that MMPI interpreters can achieve a level of consensual validity on the basis of the MMPI profile approximating the average general factor loading of the criterion judges. Certain cautions in interpreting the results were presented, especially in the light of the rather unequivocal nature of the MMPI profile used in this case.

Received April 26, 1954.

References

1. Little, K. B., & Shneidman, E. S. The validity of thematic projective technique interpretations. *J. Pers.*, in press.
2. Shneidman, E. S. (Ed.) *Thematic test analysis.* New York: Grune & Stratton, 1951.
3. Stephenson, W. S. *The study of behavior.* Chicago: Univer. of Chicago Press, 1953.
4. Tryon, R. C. *Cluster analysis.* Ann Arbor, Mich.: Edwards Brothers, 1939.

---

[3] A further study, using a variety of subjects, psychological techniques, and interpreters to increase the representativeness of the results and to test this hypothesis, is currently being conducted.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

The Validity of Taylor's Manifest Anxiety Scale [1, 2]

Edward Kendall
University of Pittsburgh

The manifest anxiety scale (*A* scale) was developed by Taylor [16] as a part of an investigation of the relation between anxiety and eyelid conditioning, in order to discriminate experimental subjects on the manifest anxiety continuum. Taylor selected approximately 200 items from the Minnesota Multiphasic Personality Inventory (MMPI) and submitted these to clinical judges with the request that they select those items which they judged to be indicative of manifest anxiety according to a definition furnished them. The 65 items on which there was 80% agreement or better were included in the scale, although in a later revision the scale was shortened to 50 scored items [14]. For a more complete description of the development and revision of this scale the reader is referred to Taylor [17]. The *A* scale has frequently been employed in investigations of anxiety and learning phenomena [1, 4, 6, 8, 10, 11, 12, 13, 15, 18], and important conclusions have been derived from these studies. Thus, a serious problem would seem to be the determination of the adequacy of the *A* scale as a measure of manifest anxiety.

The reliability of the *A* scale has been shown to vary between .81 and .96 [6, 14, 16, 17], according to the method employed, and thus it is safe to conclude that adequate reliability has been demonstrated. The outstanding deficiency in the research with the *A* scale to date is the paucity of evidence concerning its validity. As originally developed and used by Taylor the scale was not validated against any criterion of manifest anxiety external to the test itself. Taylor [17] has recently taken the position that the items of the scale may be regarded as an operational definition of manifest anxiety.

However, if anxiety is regarded as a state manifesting itself in a wide variety of behavior, and if the *A* scale is conceived of as a means of tapping such a state, then external validation is clearly required. The need for external validation is even more apparent when some of the assumptions of self-report tests are considered. Cronbach has pointed out that "the crucial assumptions involved are that the subject is willing to tell the truth to himself and to the investigator, and that the subject can determine the truth" [3, p. 307].

The currently available studies with the *A* scale provide conflicting evidence as to its validity. Rosenbaum [12], in the process of studying anxiety and stimulus generalization, found that a division of his subjects into high- and low-anxiety groups by means of the *A* scale and psychiatric ratings gave similar results. These apparently positive findings as to the scale's validity were placed in doubt by the results of Bitterman and Holtzman [1]. They found that a division of subjects by means of extensive clinical evaluation demonstrated a significant relation between anxiety and conditioning, whereas a division by means of the *A* scale alone did not produce significant findings. In a more recent study Holtzman, Calvin, and Bitterman [7] obtained *A*-scale and Winne-scale scores for a group of subjects. A correlation of .72 was obtained between the scales, which the authors interpreted as evidence for the validity of the *A* scale since the Winne scale is an empirically derived scale of neuroticism. Taylor [17] has recently presented some indirect evidence of the scale's validity. She obtained the distribution of scores for 103 neurotic and psychotic subjects, and found that the median score was equivalent to the 98.9 percentile for normal subjects. On the assumption that the former exhibit greater manifest anxiety than normals, she concluded that her findings seemed to indicate some relation between *A*-scale scores and clinical observations of manifest anxiety. However, in the same article Taylor indicated the desirability of direct validation using ratings as a criterion.

---

[1] From the Veterans Administration.

[2] This paper is based on a thesis submitted in partial fulfillment of the requirements for the degree of master of science. The writer wishes to acknowledge his indebtedness to Dr. A. W. Bendig, who originally suggested the area of investigation, and to the other members of his committee, Drs. A. D. Lazovik and H. W. Goodman. A special note of gratitude is also due Dr. Ralph Simon, Chief Psychologist at the VA Hospital, Butler, Pennsylvania, the installation at which the study was carried out.

The evidence cited as to the validity of the *A* scale is in several cases contradictory, and there is a notable lack of formal validation studies. The present study was intended as a formal validation study using a criterion of anxiety external to the test itself. The particular criterion chosen was the rating by ward nurses of the manifestly anxious behavior of chronically ill tuberculous patients.

Procedure

Development of the rating scale. A seven-point graphic rating scale was constructed for each of nine aspects of manifest anxiety. The subtraits were drawn from Cameron's description of manifest anxiety [2, p. 249], since Taylor's judges had used this definition in selecting the items of the *A* scale. The subtraits were intended to tap the extent to which the patient: (*a*) gave exaggerated and inappropriate reactions on slight provocation, (*b*) gave general indications of fatigue not attributable to his physical condition, (*c*) displayed difficulties in elimination not explainable by his physical condition, (*d*) appeared to be easily upset, (*e*) showed indications of general restlessness, (*f*) slept poorly, (*g*) displayed symptoms of nausea or vomiting not attributable to his physical state, (*h*) complained of difficulties in concentration or thinking, and (*i*) appeared to be generally tremulous. Verbal descriptions of the mid-points and end points of the scales were provided, and an example drawn from the hospital situation was provided for each of the subtraits.

A trial run of the rating scale was accomplished to ascertain its reliability and to provide information for its further refinement if this proved necessary. For this purpose, 20 randomly selected patients, 10 from each of two treatment wards, were independently rated by two nurses on each of the wards. Product-moment coefficients of the interjudge reliability were .45 and .92. Since the combined ratings by two judges were to serve as the criterion, the Spearman-Brown formula was applied and corrected values of .62 and .96 obtained. A *t* test of the difference between the two *r*'s was 2.185, significant at the .01 level. In view of this difference in reliability, an analysis was made of the ratings for the two wards, and it became apparent that the lowered reliability on the one ward could be chiefly attributed to gross disagreement in the rating of one subject. Discussion with the nurses revealed that one of them had strong negative feelings toward this particular patient, and therefore it was decided to emphasize the need for objectivity in instructing later groups of raters.

Subjects. The subjects of the study proper were 93 patients undergoing active treatment for pulmonary tuberculosis; they were randomly selected from the resident population of the VA Hospital, Butler, Pennsylvania. Their age range was from 18 to 56 years with a median of 31. The sample included 80 whites and 13 Negroes. The subjects had been given the group form of the MMPI as a part of a routine admission battery, and these tests were rescored for the items of the *A* scale. The range in obtained *A*-scale scores was from 1 to 36, with a median of 13.06, a mean of 13.15, and a standard deviation of 6.56. The distribution was somewhat positively skewed. These findings closely parallel those of Taylor [16] with a college population; she obtained a range of from 1 to 36, a median of 14, and a positively skewed distribution.

After the *A*-scale score distribution had been obtained, the upper and lower 27% were selected in order to test the ability of the *A* scale to predict manifestly anxious behavior as measured by ratings. Ratings were collected only for the 50 subjects falling in the extreme groups. Extreme groups were selected to provide a coarse test of the discriminability of the scale, corresponding to the use of the scale experimentally. The upper and lower 27 per cent maximized the chance of obtaining a significant difference [9, p. 301].

Collection of ratings. Each subject was independently rated by two nurses on the ward where he resided, and only those nurses who had observed the subjects for at least one month were used as raters. The time interval between testing and rating ranged from two to six months, with the great majority of subjects being rated within four months of testing.

In collecting the ratings the experimenter met with the two nurses from each of the wards and explained the task to them. Instruction sheets were provided, and each nurse was asked to rate a patient not included in the study in the presence of the experimenter to determine if she understood the procedure. The judges were instructed not to discuss the ratings with anyone else, and to accomplish them independently. They were allowed one week to complete the ratings.

Results

Reliability of ratings. Reliability coefficients were computed by means of the interjudge agreement method for each of the wards from which subjects were drawn. The obtained values together with the Spearman-Brown values are presented in Table 1. A chi-square test, using the z' transformation method [5, p. 135], was made to test the hypothesis that the ratings were drawn from a common population, and the obtained chi square of 6.295 with 5 *df* was not significant at the .05 level. Thus there was no basis for rejecting the null hypothesis. The average r, computed by the z' transformation [5, p. 133], was .91.

Table 1
Interjudge Reliability Coefficients of Ratings
Ward by Ward

Ward	N	r*	r_c†
1	6	.98	.99
3	7	.67	.80
4	7	.92	.96
5	12	.74	.85
6	7	.83	.91
7	11	.75	.86

* Product-moment correlation coefficients.
† Product-moment coefficients corrected by Spearman-Brown formula to provide estimate of reliability of combined ratings by two judges.

Validity of the A scale. A t test of the significance of the difference between the mean ratings of subjects falling in the upper and lower 27 per cent of the *A*-scale distribution was made. Although the mean difference was in the expected direction, the obtained t was not significant at the .05 level. However, the value of 1.407 approached significance, and in view of the suggestive nature of this finding it was decided to compute a second test of significance providing a coarser test of validity. For this comparison the upper and lower 13 per cent of the *A*-scale distribution were selected. The obtained t of 3.227 made it possible to reject the null hypothesis at better than the .01 level of confidence.

Table 2
A Comparison of the Mean Ratings of
Extreme *A*-Scale Score Groups

Group	Mean	*SD*	*SE*	t
Upper 27%	64.60	23.01	4.67	
Lower 27%	55.56	21.46	4.49	1.407*
Upper 13%	72.75	21.00	6.33	
Lower 13%	46.17	16.66	5.02	3.227**

* Not significant at the .05 level of confidence.
** Significant at the .01 level of confidence.

Discussion

Reliability of the rating scale. It was felt that the reliability of the ratings, as demonstrated, was satisfactorily high. The average reliability of .91 compares favorably with that for the *A* scale itself, for which values of from .81 to .96 have been reported [6, 14, 16, 17]. Thus an incidental result of this study would appear to be the finding that nurses who are instructed in the use of graphic rating scales are able to rate the behavior of their patients reliably. Of course this conclusion must be qualified when it is recalled that the behavior here under investigation was readily observable and that the subjects were under observation for relatively long periods of time.

Validity of the A scale. Although the original test of validity employing upper and lower 27 per cent groups yielded negative findings, a supplementary test with more extreme groups produced positive results. This procedure was felt to be justified when the nature of the t test is considered. Although the first test did not make it possible to reject the null hypothesis at the .05 level, it did not establish that no real difference existed between the two groups. The supplementary test involved increasing the chances of obtaining a large mean difference only if there were a relationship between *A*-scale scores and the ratings, and at the same time entailed a loss of degrees of freedom.

The finding of a significant difference only with very extreme groups may account for some of the differences in results in studies of the scale. A review of these studies shows that those using the preselection of extreme *A*-scale groups [13, 14, 16] have usually found a positive relationship between *A*-scale scores and simple conditioning, whereas studies not using such preselection of subjects [1, 6] have generally not demonstrated a significant relationship. The results of the present study are interpreted as supporting the validity of the *A* scale only as a coarse measure of manifest anxiety, and its experimental use is recommended only when it is desired to select extreme groups.

Summary

The present research was designed as an investigation of the validity of the *A* scale as a measure of manifest anxiety. A manifest anxiety rating scale was developed to facilitate ratings which were to serve as a criterion for validation. Ratings accomplished by means of this scale by ward nurses were found to be satisfactorily reliable when reliability was determined by the interjudge agreement method.

The *A*-scale scores for a sample of 93 hospitalized tuberculosis patients were obtained, and extreme groups were selected so as to include the upper and lower 27 per cent. Each subject was rated independently by two ward nurses, and their combined ratings served as the criterion for validation. A test of significance between the mean ratings of these two groups did not make it possible to reject the null hypothesis. A supplementary comparison between the upper and lower 13 per cent did make it possible. The results are interpreted as indicating that the *A* scale is valid only as an extremely coarse measure of manifest anxiety.

Received April 28, 1954.

References

1. Bitterman, M. E., & Holtzman, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615-623.
2. Cameron, N. *The psychology of behavior disorders.* Boston: Houghton Mifflin, 1947.
3. Cronbach, L. J. *Essentials of psychological testing.* New York: Harper, 1949.
4. Deese, J., Lazarus, R. S., & Keenan, J. Anxiety, anxiety reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46**, 55-60.
5. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
6. Hilgard, E. R., Jones, L. V., & Kaplan, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94-99.
7. Holtzman, W. H., Calvin, A. D., & Bitterman, M. E. New evidence for the validity of Taylor's manifest anxiety scale. *J. abnorm. soc. Psychol.*, 1952, **47**, 853-854.
8. Hughes, J. B., II, Sprague, J., & Bendig, A. W. Anxiety level and performance in serial learning. *J. exp. Psychol.*, in press.
9. Kelley, T. L. *Fundamentals of statistics.* Cambridge, Mass.: Harvard Univer. Press, 1947.
10. Lucas, J. D. The interactive effects of anxiety, failure and intra-serial duplication. *Amer. J. Psychol.*, 1952, **65**, 59-66.
11. Montague, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91-96.
12. Rosenbaum, G. Stimulus generalization as a function of clinical anxiety. *Amer. Psychologist*, 1950, **5**, 327. (Abstract)
13. Spence, K. W., & Farber, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116-119.
14. Spence, K. W., & Taylor, Janet A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183-188.
15. Spence, K. W., & Taylor, Janet A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265-272.
16. Taylor, Janet A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **42**, 81-92.
17. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285-290.
18. Taylor, Janet A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61-64.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

MMPI Correlates of the Taylor Scale of Manifest Anxiety[1]

Glen Brackbill
Veterans Administration, Palo Alto
and Kenneth B. Little
Stanford University

Recently a number of studies have been reported [1, 3, 5, 6, 7, 8, 11, 12] in which the subjects were selected by the use of a scale of manifest anxiety constructed by Taylor [9]. The items for this scale were selected from the Minnesota Multiphasic Personality Inventory (MMPI) as being indicative of manifest anxiety in the judgments of experts. Retest reliability figures for the scale range from .85 to .89 [10] depending in part upon the form of the test used and the interval between tests. Holtzman *et al.* [4] have reported some validity data in the form of a correlation with an independently derived scale of neuroticism by Winne [13]. Their validity figure is about .75 with some slight variation if a nonlinear index is used.

It would seem, therefore, that the scale satisfactorily measures the expression of discomfort, complaints about inefficient functioning, and general psychological malaise which can be classified under the heading of manifest anxiety.

The manifest anxiety scale (*A* scale) has been used to date primarily for selecting criterion groups in the study of the differential effects of anxiety on a variety of functions. It is obvious, however, that a scale measuring some form of anxiety could be of value to the clinician if his current tools do not permit such assessment. In interpreting the MMPI, statements about the level of anxiety of the subject are currently made on the basis of the relationship among certain of the scale scores, e.g., Depression (*D*), Psychasthenia (*Pt*) and *K*. If the *A* scale provides information not present in the scores of the MMPI scales or easily inferred from the relationships among them, then it would seem a desirable adjunct to the standard profile. The present investigation was undertaken to provide information on this point through the study of the relationship of the *A* scale to the existing scales of the MMPI.

Procedure

The product-moment correlation between the *A* scale scored from the 50 *A* items in the MMPI and each of the MMPI scales(with the exception of the Lie and the Question scales) was computed on the basis of raw scores for each of two groups.[2] The first group consisted of 106 male patients of a Veterans Administration neuropsychiatric hospital (sample 1). They constituted the entire population of usable records in the active files of the psychology service of that hospital. The modal diagnosis of the group was paranoid schizophrenia. The records of 73 male college students formed the second or "normal" group (sample 2). Scatter diagrams for all correlations were inspected and judged to be essentially linear.

Results and Discussion

Table 1 gives the correlations between the *A* scale and each of the MMPI scales for the two groups. Also indicated are the means and standard deviations of the scale score distributions.

The most striking feature of the data in Table 1 is the extremely high correlation between the Taylor anxiety scale and the MMPI *Pt* scale. In the hospital sample the

---

[1] From the VA Hospital, Palo Alto, California.

[2] We would like to express our appreciation to R. E. Billings and Yvonne Brackbill for assistance in collecting data and making the calculations.

Table 1

Correlations of the Taylor Manifest Anxiety Scale with the Scales of the MMPI for 106 Neuropsychiatric Hospital Patients and 73 College Students

Scale	Patients			Students		
		Raw score			Raw score	
	r_{Ai}	mean	*SD*	r_{Ai}	mean	*SD*
A	—	19.4	11.7	—	11.8	6.1
F	.63	10.6	7.7	.19	4.2	2.4
K	–.74	14.1	6.0	–.56	16.0	4.1
Hs	.78	10.5	8.2	.60	4.0	2.8
D	.74	26.0	8.5	.38	19.0	4.5
Hy	.50	24.4	7.5	.05	20.3	4.4
Pd	.64	21.4	6.0	.16	15.2	4.2
Mf	.45	25.6	5.5	.44	27.1	4.5
Pa	.59	13.6	5.3	.15	8.5	2.6
Pt	.92	18.9	11.9	.81	10.2	5.3
Sc	.83	22.1	14.8	.56	9.3	5.2
Ma	.28	18.0	5.3	.10	16.5	4.5

obtained correlation (.92) exceeds the highest reliabilities reported for either of the two scales, suggesting that they would make very satisfactory alternate forms. The correlation between *A* and *Pt* in the college sample (.81) is significantly lower than in the hospital group. It is possible, therefore, that among college students *A* and *Pt* are measuring something slightly different.

In order to determine whether the size of the correlation between *Pt* and *A* was the result of chance factors, the relationship between the two scales in samples from five additional clinical populations was computed.[3] These groups were: (sample 3) 60 consecutive admissions to a general medical hospital; (sample 4) 63 cases referred to the psychology section of a Veterans Administration diagnostic center; (sample 5) 60 randomly selected cases from the files of a university counseling center; (sample 6) 83 consecutive admissions to a Veterans Administration tubercular hospital; (sample 7) 59 additional neuropsychiatric hospital cases. The data thus obtained are presented in Table 2.

In each of the samples the correlation between *A* and *Pt* again exceeds the reported reliabilities of the two scales. The explanation can be advanced that the excessively high correlation is the result of correlated error terms in the two scales. They are taken simultaneously so momentary variations in mood and attitude affect each equally; both are imbedded in the same common matrix of the remainder of MMPI items inducing a similar set; and, for the 13 items the scales have in common, the identical *responses* are scored on each. Such factors will, of course, have the effect of increasing the apparent relationship between *A* and *Pt*. In terms of the expressed purpose of the study, however, the results are quite clear: with clinical populations no information would be provided by the scoring of the *A* scale in the MMPI that is not already present in the current battery of scales. It is possible that the *A* scale as administered in its own set of buffer items might provide useful additional information for the clinician since Taylor [10] found a correlation of only .68 between the two forms administered with an 18-week time interval. However, Deese *et al.* [2] report a correlation of .81 between the Biographical Inventory form of the *A* scale and the MMPI *Pt* scale in a college population when the latter was taken at the same time as the former. In a more variable clinical population it is reasonable to assume that the figure would be even higher perhaps approximating those obtained in the present study.

An additional finding of interest in Tables 1 and 2 is the remarkably low average *A* scores of the clinical samples. Taylor reports a median score of 34 for 103 neuropsychiatric subjects and, although the mean score is not given, it can be estimated as about 31 from the frequency polygon she presents [10, p. 289]. In another report by Spence and Taylor [8] the mean *A* score for 55 neurotics and psychotics is given as "about 32." In the present study the highest average *A* score of any sample is 21.9 with a standard error of the mean of 1.5. The difference between this figure and those obtained by Taylor, and Spence and Taylor, is obviously quite significant but no ready explanation is apparent for the discrepancy.

The comparison of the average MMPI profiles for the high and low scores in both groups gives some suggestion as to the meaning of the

---

[3] The authors are indebted to Dr. Jerome Fisher, VA Hospital, San Francisco; Dr. Roger Bardsley, VA Hospital, Livermore; and Dr. John Black, Counseling Center, Stanford University, for their assistance in compiling these data.

Table 2

Correlations Between the Taylor Manifest Anxiety Scale and the MMPI *Pt* Scale in Five Samples of Clinical Populations

Sample	Population	*N*	$r_{A\,Pt}$	M_A	SD_A	M_{Pt}	SD_{Pt}
3	General medical patients	60	.92	16.1	8.7	11.1	8.1
4	Diagnostic center patients	63	.91	16.1	9.2	12.4	9.6
5	Counseling center clients	60	.93	12.3	7.7	11.4	7.2
6	TB hospital patients	83	.91	14.6	8.7	13.1	8.2
7	NP hospital patients	59	.92	21.9	11.6	18.5	11.6

A scores in different populations. Table 3 presents the mean raw scores for the highest and lowest 10 subjects of the *A*-score distributions of the college and neuropsychiatric hospital groups.[4]

Table 3

Average MMPI Scale Scores of the Ten Highest and Ten Lowest *A*-Scale Cases of the Student and Hospital Groups

Group	L	F	K	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma	A
College Low A	4	3	21	2	16	21	14	22	8	3	4	4	3
College High A	4	5	12	7	23	21	15	29	8	17	13	15	22
Hospital Low A	6	2	21	9	19	20	17	22	9	4	6	16	3
Hospital High A	3	17	10	21	36	31	26	28	17	33	38	21	39

For the high scorers in the hospital, the profile is an obviously psychotic one although the relation of *F* and *K* suggests that there is some tendency for the subjects to place themselves in the worst possible light. Overt anxiety and turmoil are present as suggested by the high *D* score and the profile in general suggests that the patients recognize their illness and are actively seeking help.

The mean profile of the low *A* scorers of the hospital group, on the other hand, suggests more of the character disorder. The existence of denied anxiety is implied by the elevated *K* score, but the acting out of impulses indicated by the peak on *Pd* reduces much of the tension that might produce manifest anxiety symptoms.

In the college group, the higher scorers' average profile suggests that they are introspective, quite sensitive to environmental press, and willing to admit to being easily disturbed. The low scorers, conversely, utilize both denial and repression fairly frequently and rarely introspect.

Summary

The relationship between the Taylor Manifest Anxiety Scale, as scored in the MMPI, and the other MMPI scales was examined in an attempt to discover what additional information would be provided by its use. Two samples, a college student group and a neuropsychiatric hospital group, were used. On the basis of the extremely high correlation between the *A* scale and the MMPI *Pt* scale (.92), it was concluded that manifest anxiety could be ascertained from the latter in clinical populations as reliably as from the former. The meaning of high and low scores on the *A* scale of the two groups was discussed in terms of mean MMPI profiles of extreme groups.

Received April 13, 1954.

References

1. Bitterman, M. E., & Kniffin, C. W. Manifest anxiety and "perceptual defense." *J. abnorm. soc. Psychol.*, 1953, **48**, 248-252.
2. Deese, J., Lazarus, R. G., & Keenan, J. Anxiety, anxiety reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46**, 55-60.
3. Hilgard, E. R., Jones, L. V., & Kaplan, S. J. Conditioned discrimination, as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 44-99.
4. Holtzman, W. H., Calvin, A. D., & Bitterman, M. E. New evidence for the validity of Taylor's manifest anxiety scale. *J. abnorm. soc. Psychol.*, 1952, **47**, 853-854.
5. Montague, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91-96.
6. Spence, K. W., & Farber, I. E. The relation of anxiety to differential eyelid conditioning. *J. exp. Psychol.*, 1954, **47**, 127-134.
7. Spence, K. W., & Taylor, Janet A. Anxiety and strength of the UCS as determiners of the

---

[4] To conserve space, *T*-score profiles are not presented. The interested reader may plot them on the standard MMPI scoring sheet using the data of Table 3.

amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183-188.

8. Spence, K. W., & Taylor, Janet A. The relation of conditioned response strength to anxiety in normal, neurotic and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265-272.
9. Taylor, Janet A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81-92.
10. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285-290.
11. Taylor, Janet A., & Spence, K. W. The relationship of anxiety to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61-64.
12. Wesley, Elizabeth L. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.*, 1953, **48**, 129-134.
13. Winne, J. F. A scale of neuroticism: an adaptation of the Minnesota Multiphasic Personality Inventory, *J. clin. Psychol.*, 1951, **7**, 117-122.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

Regional Differences in MMPI Responses Among Male College Students[1]

Leonard D. Goodstein[2]
State University of Iowa

The usefulness of any psychometric instrument depends, at least in part, upon the applicability of the available norms. The blind application of a test valid for one population to a different type of group often leads to serious errors of interpretation. The use of the MMPI with college populations has been criticized on two points. First, college students tend to be more deviant than the general population in their MMPI responses and, second, local or regional norms are thought necessary for each college or geographical region. Brown [2], Cottle [4], Gilliland and Colgin [7], and others have urged the development of local norms while Tyler and Michaelis[3] among others have actually constructed such local norms.

Black [1] has taken a somewhat contradictory position. He collected, from the literature, MMPI data on a large group of college females, a total N of 5014, at 15 different colleges and universities. Although he did not use an over-all test of significance, he concluded, "these data suggest that there is a characteristic profile for college women which does not differ from college to college. It is certainly true that some of the differences are statistically significant, but that they are of little practical significance" [1, p. 34]. The purpose of the present paper is further to test the notion that local or regional norms are necessary for interpreting MMPI profiles of college students by examining the data available on college males.

Procedure

One group of *Ss*, consisting of 408 randomly selected freshman males, had been tested at the State University of Iowa using the booklet form of the MMPI. From the published literature and from those unpublished studies available to the author, seven additional groups were selected as meeting the following criteria:

a. *Ss* randomly selected or 100 per cent sampling technique used. The inclusion of data that had been obtained through administration to students in special curricula or to preselected groups of students, e.g., student leaders, might lead to a confounding of the results and were not included.

b. $N > 100$. Relatively large samples were seemingly more free of bias.

c. Means and standard deviations reported for all nine clinical scales with *K* correction where appropriate. As the *K* correction is now routinely used, those studies without the *K* correction were not included.

Results

The means, the standard deviations, and the number of cases on the nine clinical scales and the *K* scale for the eight schools are presented in Table 1. The agreement among the scale means of the eight samples is very striking. The largest difference between any two means on a single scale is 7.8 *T*-score units on the *Mf* scale. With the exception of the differences on the *Mf* scale, none of the differences exceeds 4.9 *T*-score units, which would be less than one-half of a standard score according to the original standardization data. The smallness of the differences using raw scores is even more striking. Only the *Mf* scale has a difference between any two means of over 4 raw score units. The median range difference in raw score units is only 1.9.

The data from these eight colleges were then placed into three groups based upon geographical regions: the Eastern group, including

---

[1] Based on a paper presented at the annual meeting of the Midwestern Psychological Association, Columbus, Ohio, April 30, 1954.

[2] The author wishes to acknowledge his indebtedness to Robert A. Johnston for his assistance with the statistical analysis of these data.

[3] F. T. Tyler & J. U. Michaelis. University student norms for the Minnesota Multiphasic Personality Inventory. Unpublished manuscript, School of Education, University of California, Berkeley.

Table 1

MMPI Means and Standard Deviations in *T* Scores for Each College Arranged According to Geographical Region

College	*N*		*K*	*Hs**	*D*	*Hy*	*Pd**	*Mf*	*Pa*	*Pt**	*Sc**	*Ma**
East												
Maine [10]	316	M	†	53.0	53.0	55.0	55.0	59.0	50.0	56.0	55.0	60.0
		SD	†	8.5	12.1	7.7	10.8	10.6	9.1	12.6	12.2	11.1
Penn. State [3]	121	M	†	53.4	53.5	56.3	57.0	59.6	52.2	58.0	57.1	60.2
		SD	†	8.3	11.8	7.7	9.5	8.6	9.6	10.8	8.9	9.6
Midwest												
Iowa	408	M	15.3	53.0	53.0	56.0	57.0	58.0	54.0	56.0	57.0	58.0
		SD	4.8	8.2	11.6	8.2	9.8	10.8	7.9	12.8	11.6	11.0
Minnesota [11]	1321	M	14.4	51.1	51.8	54.2	56.6	56.9	52.6	56.8	56.9	58.1
		SD	4.6	8.1	10.3	7.3	9.7	9.7	8.6	10.4	10.8	10.3
Wisconsin‡	1422	M	14.5	51.2	52.2	55.0	56.0	59.0	53.0	56.7	56.0	58.0
		SD	4.4	7.2	10.5	7.7	9.7	10.6	7.9	9.4	10.1	10.6
West												
Montana State‡	456	M	14.2	50.5	52.5	54.5	56.0	58.0	54.1	56.9	56.9	59.0
		SD	4.5	10.0	12.1	7.8	9.6	10.2	8.0	10.1	8.9	9.6
New Mexico [9]	149	M	16.4	54.1	53.9	57.8	58.1	62.7	53.4	56.8	57.2	59.2
		SD	9.4	8.7	10.5	7.8	10.4	9.9	7.8	10.2	10.5	10.0
Utah State [5]	842	M	13.8	51.5	50.2	52.9	54.3	54.9	53.0	53.6	52.4	55.8
		SD	4.4	7.1	9.0	7.9	8.8	9.1	9.0	9.5	10.7	9.2
Median	5035	M	14.5	52.3	52.8	55.0	56.3	58.5	53.0	56.7	56.9	58.7
		SD	4.6	8.3	11.1	7.8	9.8	10.1	8.3	10.3	10.8	10.2

* Includes K correction.

† K added but its value not reported.

‡ From an unpublished paper by Donald P. Hoyt, Student Counseling Bureau, University of Minnesota.

The present author wishes to thank Dr. L. E. Drake and Dr. G. A. Renzaglia for permission to publish the Wisconsin and Montana State data.

the University of Maine and the Pennsylvania State College; the Midwestern group, including the Universities of Iowa, Minnesota, and Wisconsin; and the Western group, including Montana State College, Utah State College, and the University of New Mexico. To provide an over-all test of significance these data were analyzed by a Lindquist [8] Type I analysis of variance design in which the regions were a "between *S*s" factor, while the nine clinical scales and the regions × scales interaction were "within *S*s" factors. As shown in Table 2, only the obtained *F* for the scales was significant, indicating that the means on the nine clinical scales were not equal. The obtained *F*'s for regions and the regions × scales interaction did not exceed unity. There is no evidence to support the notion that geographical differences are significant determinants of MMPI means.

Table 2

Analysis of Variance of Means

Source	*df*	MS	*F*
Between *S*s	7	11.37	
Regions	2	1.69	
Error (*b*)	5	15.25	
Within *S*s	64	67.55	
Scales	8	46.71	56.72**
Regions × Scales	16	1.45	1.63
Error (*w*)	40	.89	
Total	71		

** p < .01.

Table 3

Analysis of Variance of *SD*'s

Source	*df*	MS	*F*
Between *S*s	7	2.34	
Regions	2	2.13	
Error (*b*)	5	2.43	
Within *S*s	64	1.87	
Scales	8	10.81	18.02**
Regions × Scales	16	.57	
Error (*w*)	40	.60	
Total	71		

**p < .01

To test the importance of these geographical regional influences, a similar analysis was then performed with the standard deviation data. The variability data were also divided into three groups in the same manner as the mean data and the identical analysis of variance design was used.[4]

As may be seen in Table 3, only the obtained F for scales was statistically significant, which would indicate that the variability of the nine clinical scales was not equal. The D and Pt scales have the largest variances while the Hy, Hs, and Pa scales have the smallest variances. The F's for regions and the regions × scales interaction were less than unity. On the basis of this analysis, therefore, we may conclude that the variances are not significantly dependent upon regional influences.

Discussion

An inspection of the means for the eight schools reveals that no one group was consistently higher or lower than any other group. The patterning of scores would seem to be more a function of chance fluctuation than due to the operation of any cultural or environmental factors. The obtained differences over regions would seem to be of such little consequence that the development of regional norms seems unnecessary.

It is important to note that the mean T scores on all nine clinical scales are above the expected mean value of 50. This finding is in support of the notion that college students, as a group, are more deviant in their responses to the MMPI than the general adult population used in the standardization of the instrument. These results, however, should not be interpreted to mean that the MMPI cannot be useful in evaluating the adjustment of college students, but rather supports the idea that separate norms for college students as a group are not only desirable but essential.

In this regard it is interesting to note that, using 70 as the cutting score, approximately 15% of our total group of 5000, or 750, have "abnormal" scores on the Ma scale. Extending this criterion to the other scales, we identify approximately 7 to 10% of our group as abnormal on each of the Pd, Mf, Pt, and Sc scales. Without allowing for any overlap, that is, a single S scoring above 70 on more than one scale, we would have identified approximately half of our 5000 male college students as abnormal by such a procedure. While this, of course, may be a valid "diagnosis," the capacity of our college mental hygiene facilities certainly demands a more rigorous screening instrument. Seemingly the usefulness of the MMPI as a screening test in the collegiate setting would depend upon the development of new cutting scores.

The consistency in the pattern of scores among the eight schools also seems worthy of note. The peaks on Ma, Mf, Sc, Pt, Pd, and Hy suggest that there is a characteristic profile for college males. They appear to be more feminine in their interests, to be more active, less inhibited, but more worrying than the male population in general. This diagnostic picture would seem to be validated by clinical observations in the classroom and on the campus as well as in the counseling office.

We may then compare our typical profile for college males with that obtained by Black [1] for college females. This comparison is shown in Figure 1. In each case the points are based upon the median mean, in our study the median is of 8 colleges, in Black's study it is of 15 colleges. The total N's, however, are very similar, 5035 for our sample, 5014 for Black's.

The differences between these two profiles are rather interesting. The males are higher on all the clinical scales with the exception of Pa where the points almost coincide. The largest differences are on the Mf, Ma, Pt, D, and Hs scales. The women are below the expected mean value of 50 on the Hs and D scales while the men, as noted before, are consistently above 50. The women are higher on the K scale, indicating a somewhat more defensive attitude. The effect of this higher K value, however, is to reduce the differences between the sexes on K-corrected scales as the higher K results in higher values on five of the nine scales.

While it may not be obvious upon first glance, there is a good bit of similarity between these two profiles. If we exclude the Mf and Pa scales from consideration, the profiles then appear quite similar with peaks on Ma and Sc and valleys on Hs and D, although the fe-

---

[4] While such data do not usually fulfill the assumptions about normality necessary for an analysis of variance, recent empirical evidence would indicate that slight deviations from normality are of little consequence and may be ignored [8, pp. 78-90].

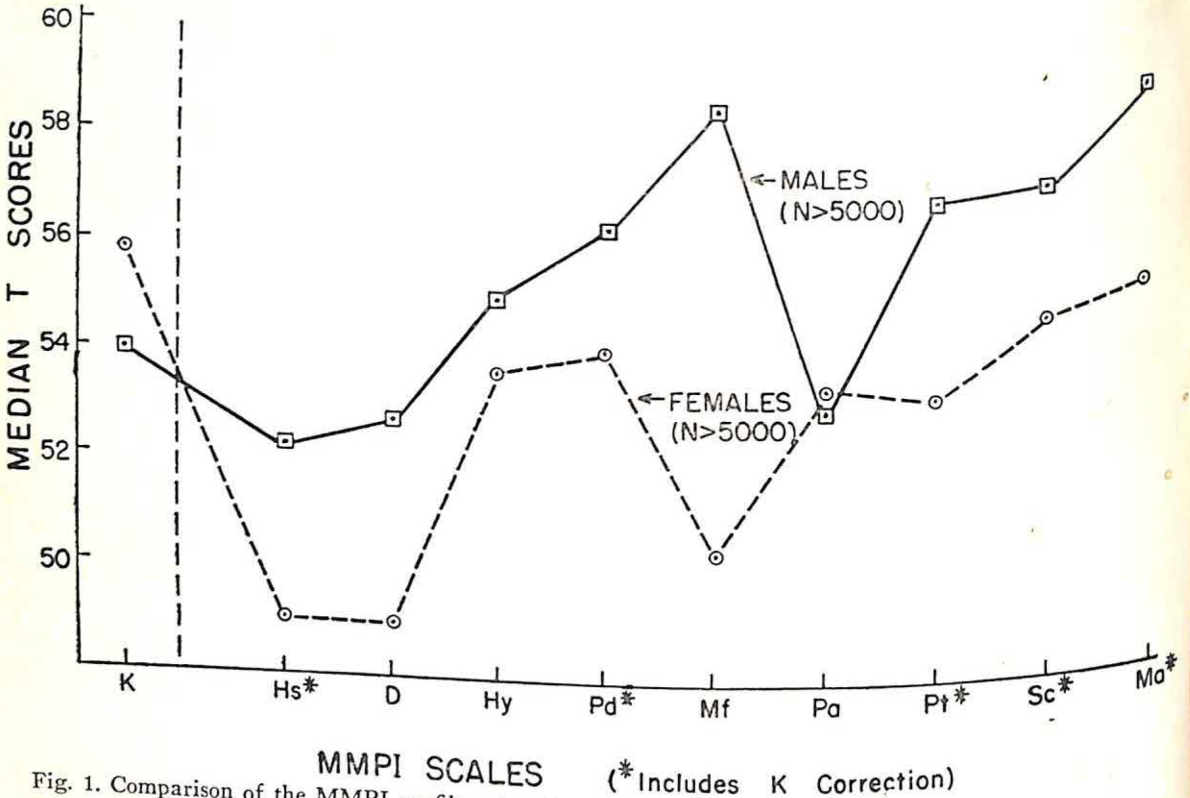


Fig. 1. Comparison of the MMPI profiles of male and female college students.

male profile is, of course, lower. It is rather surprising to note that, while college males are considerably more feminine in their interests than males in general, college females are *not* more masculine in their interests than women in general. The reasons for this finding and the possible implications for the future adjustment of these individuals are left to conjecture. The differences between these profiles would suggest that, in using the MMPI in a college setting, separate norms are necessary for males and females.

Summary

The present study has been a comparison of the results of MMPI administration to large groups of male undergraduates at eight different colleges and universities for regional differences. When the mean scores were subjected to an analysis of variance, the results indicated that there were no significant regional differences. An analysis of variance for the standard deviations also showed no significant regional differences. The obtained differences would seem to be of such little consequence that the development of regional or local norms seems unnecessary.

While there is a characteristic profile for the college male that differs little from college to college, it is markedly different from the characteristic profile of the noncollege male and from the characteristic profile of the college female. New norms would appear necessary in using the MMPI as a screening test in clinical work with university populations.

Received May 10, 1954.

References

1. Black, J. D. The interpretation of MMPI profiles for college women. Unpublished doctor's dissertation, Univer. of Minnesota, 1953.
2. Brown, H. S. Similarities and differences in college populations on the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1948, **32**, 541-549.
3. Cauffiel, P. W., & Snyder, W. W. A comparison of the performance of a randomly selected college population on the MMPI and the P-S Experience Blank. *J. clin. Psychol.*, 1951, **7**, 267-270.
4. Cottle, W. C. The MMPI; a review. *Kans. Stud. Educ.*, 1953, **3**, No. 2.
5. Dobson, W. R., & Stone, D. R. College freshman responses on the MMPI. *J. educ. Res.*, 1951, **44**, 611-618.

6. Fry, F. D. A study of the personality traits of college students and of state prison inmates as measured by the MMPI. *J. Psychol.,* 1949, **28**, 439-449.

7. Gilliland, A. R., & Colgin, R. Norms, reliability and forms of the MMPI. *J. consult. Psychol.,* 1951, **15**, 435-438.

8. Lindquist, E. L. *Design and analysis of experiments.* Boston: Houghton Mifflin, 1953.

9. Norman, R. D., & Redlo, M. MMPI personality patterns for college major groups. *J. appl. Psychol.,* 1952, **36**, 404-409.

10. Sopchak, A. L. College student norms for the Minnesota Multiphasic Personality Inventory. *J. consult. Psychol.,* 1952, **16**, 445-448.

11. Williamson, E. G., & Hoyt, D. Measured personality characteristics of student leaders. *Educ. psychol. Measmt,* 1952, **12**, 65-78.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

Social Desirability as a Factor in Edwards' Personality Preference Schedule Performance[1]

Leslie Navran and James C. Stauffacher
VA Hospital, American Lake, Washington

The Edwards Personality Preference Schedule (EPPS) is a new personality inventory which yields scores for 15 of the needs in Murray's theoretical system. These are: achievement, deference, conjunctivity, exhibition, autonomy, affiliation, intraception, succorance, dominance, abasement, nurturance, change, endurance, heterosexuality, and aggression. In constructing the schedule, Edwards has utilized the knowledge that there is a linear relationship between the probability of endorsement of an item and its judged social desirability ($r = .871$). Thus, the structure of the instrument is such that for each of the 225 items, the subject has to choose between two statements which are matched for social desirability, designating his choice as being more characteristic of him.

This paper presents data from a project designed to evaluate the usefulness of the EPPS for counseling and guidance in the field of nursing.

The EPPS was administered to 25 student affiliate nurses spending three months in residence at VA Hospital, American Lake, Washington. A month later, the girls separately ranked cards which bore the names and definitions of the needs sampled in the EPPS so that they ranged from most to least characteristic of themselves. Then, they ranked the 15 needs again from most to least socially desirable.

Composite ranks were made of the 15 needs from the EPPS raw scores, from the students' self-descriptions, and finally from their judgments of social desirability. The composite EPPS raw-score ranking correlated −.03 with the self-description composite, and −.01 with the social desirability composite. However, the self-description rankings correlated .90 with social desirability judgments! An analysis was made of the same correlations for each individual. While none of the 25 correlations between EPPS scores and social desirability judgments was significantly different from zero, 15 of the correlations between self-descriptions and social desirability were significantly different from zero at the .05 level ($\rho = .44$ or more). At the .01 level, 13 correlations were above the required value of .59 (one-tailed test).[2]

These results lend impressive support to the proposition that the EPPS not only has eliminated the influence of social desirability at the item level, but has also considerably reduced its operation at the variable level. Furthermore, it is felt that the criterion of self-description has been shown to be so heavily influenced by social desirability considerations as to be useless. More objective criteria are needed to attack fruitfully the issue of validity.

---

[1] An extended report of this study may be obtained without charge from Leslie Navran, VA Hospital, American Lake, Washington, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4395 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[2] This study has since been replicated with a second group of 27 affiliate nurses, with almost identical results.

Brief Report
Received July 20, 1954.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

The Relation Between Acceptance of Self and Acceptance of Others Shown by Three Personality Inventories[1]

Katharine T. Omwake
Agnes Scott College

In recent years psychiatrists and clinical psychologists have observed a relation between the attitude toward the self shown by the patient and his attitude toward other people. Adler [1] noted a depreciation of others in those who themselves felt inferior. Horney [5] asserted that the person who does not love himself is incapable of loving others. Rogers [8] said that the person who accepts himself will for that very reason have better interpersonal relations with others. He also observed that, during therapy, as a person begins to accept himself he becomes capable of experiencing this attitude toward others. Changes in acceptance of the self and correlated changes in the acceptance of others occurring during client-centered therapy were studied by Sheerer [9]. Her study, and a similar one by Stock [10] also based on a small number of counseling cases, showed that perceptions of others, feelings toward others, and acceptance of others are significantly related to the perception of the self, feelings about the self, and acceptance of the self. These studies show that, in the course of therapy, with the increase in self-acceptance there is a corresponding increase in favorable attitude toward others.

Several attempts have been made to see to what extent these observations made by clinicians hold true for larger, more normal populations. Phillips [6] constructed a questionnaire on Attitudes Toward the Self and Others which he administered to several groups of students. In a university class composed chiefly of older students the correlation between attitudes toward the self and those toward others was .74, while with a younger group of college age it was .54. Berger [2] used the definitions of acceptance of self and acceptance of others which Sheerer [9] had made as the basis for the development of a questionnaire or scale. He found a correlation of .65 between acceptance of self and acceptance of others for evening students and of .36 for day students. This is in agreement with the study by Phillips [6], who also found a closer relation between acceptance of self and acceptance of others in the older group of college students than in the younger group. Bills, Vance, and McLean [4] devised an Index for Adjustment and Values which measures self-acceptance, and a corresponding form which shows the individual's perception of how other people accept themselves.

Phillips [6], Berger [2], and Bills, Vance, and McLean [4] have published studies of the reliability and validity of their tests, and the results obtained from administration of them to college students and to some other groups, but they have not published the tests themselves. In the present study all three of these unpublished tests, by permission of their authors, have been administered to a group of over one hundred college students.

Problem

This study is designed to test the assumption that, in a normal population, there is a positive relation between the acceptance of self and the acceptance of others, that those who are self-acceptant are also acceptant of others, and that those who reject themselves also tend to reject others.

A second hypothesis is that there should be agreement among tests designed to measure the same trait; that, therefore, various tests of acceptance of self should give similar results, and that various measures of acceptance of others, likewise, should agree.

---

[1] This article is an abbreviated version of a paper read at the meeting of the Southern Society for Philosophy and Psychology in Atlanta, April 1954.

Method

Three unpublished personality inventories, made available to the investigator by their authors, were used in this study. They are: the scale for Self-Acceptance and Acceptance of Others by Berger [2], the questionnaire on Attitudes Toward the Self and Others by Phillips [6], and the Index of Adjustment and Values for self and for others, by Bills, Vance, and McLean [4]. Each was mimeographed without a title, so there was nothing to indicate what the inventory was designed to measure. These three personality inventories, described below, were administered to 113 students in the first course in psychology at Agnes Scott College. The students were told that the investigator wished to make a study of the personality tests, and that in order to make the study it was not necessary to know the identity of those taking the tests. Each student was assigned a number which she placed on each test. It was hoped that this method of preserving anonymity would increase frankness in answering the questionnaire.

The scale for measurement of Self-Acceptance and Acceptance of Others by Berger consists of 36 items concerned with acceptance of self and 28 with the acceptance of others, mixed in random order. The following is typical of the statements showing attitude toward the self: "I realize that I'm not living very effectively but I just don't believe I've got it in me to use my energies in better ways." "I believe that people should get credit for their accomplishments, but I very seldom come across work that deserves praise" shows attitude toward others. The person taking the test is to rate each item on a five-point scale.

The Attitudes Toward the Self and Others questionnaire by Phillips is basically similar to Berger's scale. It consists of 25 statements showing attitudes about the self and 25 showing attitudes toward others, mixed in random order. The person taking the test is to rate attitude on each statement on a five-point scale.

The Index of Adjustment and Values by Bills, Vance, and McLean consists of 49 adjectives, such as *acceptable, busy, calm, poised,* and *tactful*. The subject is asked to use each word to complete the sentence "I am a (an) .... person," using a five-point scale to indicate how much of the time this is like him, and then to indicate, by rating, how he feels about himself as described in the first rating. The sum of these second ratings gives the measure of self-acceptance used in this study. The Index for others is similar. The subject uses the same 49 adjectives in turn to complete the statement "Other people are .... persons," with friends particularly in mind, and indicates how other people feel about themselves as they are, using a five-point scale as before. This gives a measure of the extent to which others are seen as accepting themselves.

Results and Discussion

The three inventories measuring self-acceptance agree markedly, as shown by the correlations in Table 1. As might be expected, there is closest agreement between the tests which are most similar in form and content. Degree of acceptance of self given by Berger's scale correlates .73 with degree of self-acceptance in Phillips' scale. Bills's Index, which measures self-acceptance by a different technique, is in less close agreement with the other two scales, although even here the correlation of .49 with Berger's scale and .55 with Phillips' indicates a substantial correspondence. All three correlations are significant at the 1 per cent level of confidence.

Table 1

Correlations of Measures of Acceptance of Self and Acceptance of Others

Tests	Correlation	Level of significance
Acceptance of self vs. acceptance of self		
Berger vs. Phillips	.73	1%
Bills vs. Berger	.49	1%
Phillips vs. Bills	.55	1%
Acceptance of others vs. acceptance of others		
Berger vs. Phillips	.60	1%
Bills vs. Berger	.23	5%
Phillips vs. Bills	.13	–
Acceptance of self vs. acceptance of others		
Berger vs. Berger	.37	1%
Bills vs. Bills	.39	1%
Phillips vs. Phillips	.41	1%
Berger vs. Phillips	.25	5%
Berger vs. Bills	.23	5%
Bills vs. Berger	.23	5%
Bills vs. Phillips	.18	–
Phillips vs. Bills	.30	1%
Phillips vs. Berger	.34	1%

There is slightly less agreement among the tests showing attitudes toward others. Here also the two tests which are similar in construction agree most closely. The correlation between the tests by Berger and Phillips is .60, significant far beyond the 1 per cent level, but lower than the correlation of .73 between the tests of acceptance of self by the same authors. Since Bills's Index for others shows the individual's perception of how other people accept themselves, rather than the degree to which he is acceptant of others, it is understandable that it correlates less well with tests measuring the individual's attitude toward others. Bills's Index correlates .23 with Berger's scale, significant only at the 5 per cent level, while the correlation of .13 with Phillips' scale is not statistically significant.

There is a consistent tendency for those who accept themselves to be acceptant of others and to view others as being self-acceptant, and, on the other hand, for those who have a low opinion of themselves to reject others also, and to see others as rejecting themselves. This is seen in the correlations of .37, .39, and .41, significant at the 1 per cent level, between the acceptance of self and attitudes toward others as measured by the inventories by Berger, Bills, and Phillips.

These correlations are similar to those obtained by Berger [2] and by Phillips [6] in groups of the same age as the group in this study. Berger's correlation is .37 as compared with .36 in the present study; Phillips' correlation of .54 shows a closer relationship than does the correlation of .41 obtained with the present group. Bills has not reported the correlation between self-other attitudes, so no comparison is possible.

As might be expected, each test of self-acceptance agrees more closely with the corresponding test of attitudes toward others made by the same author than it does with the scale for others made by another author. In every instance the correlation between the self-other parts of the same test is significant at the 1 per cent level of confidence, whereas Phillips' test of self-acceptance is the only one which gives correlations of this degree of statistical significance with tests of attitudes toward others made by another investigator. Three of the four other correlations between self-attitudes measured by one test and attitudes toward others obtained from another test are significant at the 5 per cent level. As shown by the correlations in Table 1, there is a consistent tendency for self-attitudes to be reflected in attitudes toward others.

A comparison was made of the attitude toward others held by those who represented the extremes of self-acceptance and of self-rejection. The self-acceptant group consisted of 19 subjects who were in the upper three deciles on all three tests of self-acceptance; the group who rejected themselves was composed of 12 individuals who were consistently in the lowest three deciles on all three tests. The larger number in the self-accepting group shows that a high degree of self-acceptance is more consistently shown than is a low degree; only a small number are extremely self-rejectant on all three tests, while a larger number consistently hold themselves in high esteem. On each test the mean score on attitude toward others made by the self-acceptant group is higher than the mean score made by the self-rejecting group. The difference between the means is significantly different from zero well beyond the 1 per cent level of confidence. This difference between the means on the tests by Berger and by Phillips indicates that those who are most self-acceptant also have a high degree of acceptance of others, while those who reject themselves also hold others in low esteem. The difference between the means on Bills's Index would indicate that those who are self-acceptant perceive others as accepting themselves, while those who are least self-acceptant perceive others as rejecting themselves. As had been assumed, there is a direct relation between the attitude an individual holds toward himself and that held toward others.

The data presented here, although based on personality inventories that do not probe very deeply into personality dynamics, verify in a fairly large group of college students the conclusions derived by clinicians from therapy sessions. There is evidence that in the normal population, as well as in those undergoing therapy, attitudes toward the self appear to be reflected in attitudes toward other people: the lower the opinion of the self, the lower the opinion of others. Only when the self is regarded with a fairly high degree of acceptance is it possible to relate effectively to others, to understand them, and to regard them as persons of worth.

Summary

To test the hypothesis that in a normal population there is a positive relation between acceptance of self and acceptance of others, three unpublished tests measuring attitudes toward the self and toward others were administered to 113 college students who took them anonymously. The tests are: the scale for Self-Acceptance and Acceptance of Others by Berger, the questionnaire on Attitudes Toward the Self and Others by Phillips, and the Index of Adjustment and Values by Bills, Vance, and McLean. The three measures of self-acceptance agree closely; those for attitudes toward others agree less well. The results support the hypothesis in that there is a marked relation between the way an individual sees himself and the way he sees others; those who accept themselves tend to be acceptant of others and to perceive others as accepting themselves; those who reject themselves hold a correspondingly low opinion of others, and perceive others as being self-rejectant.

Received March 30, 1954.

References

1. Adler, A. *The neurotic constitution.* New York: Dodd, Mead, 1926.
2. Berger, E. M. The relation between expressed acceptance of self and expressed acceptance of others. *J. abnorm. soc. Psychol.*, 1952, **47**, 778-782.
3. Bills, R. E. A validation of changes in scores on the Index of Adjustment and Values as measures of changes in emotionality. *J. consult. Psychol.*, 1953, **17**, 135-138.
4. Bills, R. E., Vance, E. L., & McLean, O. S. An Index of Adjustment and Values. *J. consult. Psychol.*, 1951, **15**, 257-261.
5. Horney, Karen. *New ways in psychoanalysis.* New York: Norton, 1939.
6. Phillips, E. L. Attitudes toward self and others: a brief questionnaire report. *J. consult. Psychol.*, 1951, **15**, 79-81.
7. Roberts, G. E. A study of the validity of the Index of Adjustment and Values. *J. consult. Psychol.*, 1952, **16**, 302-304.
8. Rogers, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.
9. Sheerer, Elizabeth T. An analysis of the relationship between acceptance of and respect for self and acceptance of and respect for others in ten counseling cases. *J. consult. Psychol.*, 1949, **13**, 169-175.
10. Stock, Dorothy. An investigation into the interrelations between the self-concept and feelings directed toward other persons and groups. *J. consult. Psychol.*, 1949, **13**, 176-180.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

Self-Acceptance and Its Relation to Conflict

Herbert Zimmer
Air Force Personnel and Training Research Center

Raimy [4] was able to show a shift in self-evaluation in successfully counseled clients, which did not occur in unsuccessfully counseled clients. The shift was one from self-disapproval to self-approval. Following his lead Bills, Vance, and McLean [2] defined personal maladjustment "as any discrepancy between the concept of self and the concept of the ideal self." Their Index of Adjustment and Values uses the total of the discrepancies between the self-concept and the concept of the ideal self as a measure of adjustment. Roberts [5] demonstrated significantly longer reaction times on a word association test for words with self-concept–ideal-self discrepancies than for words without discrepancies, and considered such discrepancies a "valid index of emotionality." Bills [1], upon repeating Roberts' experiment with the identical method, found a trend in the direction of Roberts' results.

The present study was undertaken to check the efficacy of self-concept–ideal-self discrepancies as indicators of conflict, and by inference of maladjustment. It tested the hypothesis that the presence of conflict over a personality trait is associated with a self-concept–ideal-self discrepancy on that trait. A confirmation of this hypothesis would support Roberts' findings [5], and provide a direct rationale for the use of instruments employing self-concept–ideal-self discrepancies as a measure of adjustment, and of improvement in counseling situations. It should be noted that this experiment was carried out prior to the publication of Roberts' [5] and Bills' [1] articles, and has therefore approached the same problem independently and by means of a different method.

Procedure

Fifty-two subjects rated themselves (*a*) as they are, and (*b*) as they would like to be with respect to 25 personality traits on two similar seven-point rating scales. By using the trait adjectives as stimulus words in a word association and reproduction test, six of the traditional complex indicators were obtained, and employed as an index of conflict.

The subjects were 26 male undergraduate students at the University of Rochester and 26 male neuropsychiatric patients at the Canandaigua, New York, Veterans Administration Hospital. A specially constructed 20-item vocabulary test was used to screen all subjects with an inadequate understanding of words of the level of difficulty used in the experiment.

The 25 trait adjectives selected for use in the experiment were: leisurely, emotional, ambitious, deliberate, trusting, sentimental, cautious, refined, energetic, orderly, poetic, wary, lusty, dominant, spontaneous, economical, respectful, daring, precise, meek, persistent, ardent, conventional, determined, and obedient. They were chosen on the basis of (*a*) occurring 6 to 10 times in one million words in the Thorndike-Lorge word count [7] column G, (*b*) not representing a cultural stereotype of a desirable or undesirable personality trait, and (*c*) having personal and dynamic meaning. Word frequency was controlled because of its possible effect on word association reaction time. Stereotyped traits were eliminated, since consistently extreme ratings might affect the size of self-concept–ideal-self discrepancies. The order of presentation of stimuli was varied systematically to counteract position effects.

Each trait was evaluated by the subject on two seven-point rating scales. The self-concept scale ran from "*7.* I am a very . . . person" to "*1.* I am definitely not a . . . person." The ideal-self scale ran from "*7.* I would like to be a very . . . person." to "*1.* I definitely don't want to be a . . . person." The complete list of traits was first rated on one scale, and later on the other scale.

The complex indicators in the word association test, which were used as the index of conflict, were (*a*) long reaction time, (*b*) long reproduction time, (*c*) defective reproduction, (*d*) repetition of the stimulus word, (*e*) responding with more than one word, and (*f*) obvious and clear-cut manifestations of overt emotional behavior. Reaction time was recorded with a Standard Electric Timer. Three of the indicators are identical to the complex indicators in Hull and Lugoff's [3] list which show the highest positive correlation with the combination of all the complex indicators used by them. Each of the six indicators was weighted equally, since no empirical system of assigning differential weights, which would designate the relative merit of the various indicators in signifying conflict, has yet been devised.

Table 1 indicates that traits with zero complex indicators, as well as traits which evoked two or more complex indicators, differentiated successfully between two groups of distinctly different emotional adjustment, while traits with one complex indicator did not. For this reason a nonconflictual trait was defined as one which did not elicit any complex indicators, and a conflictual trait as one which evoked two or more complex indicators.

Table 1

Differences between Hospital and College Groups in Mean Number of Traits Eliciting Complex Indicators

Statistic	Subjects: Hospital group ($N = 26$)	Subjects: College group ($N = 26$)	Difference
	Traits with 0 indicators		
Mean	6.23	8.27	2.04
SD	2.19	1.93	
t			3.505
p			.001
	Traits with 1 indicator		
Mean	5.69	6.23	.54
SD	2.18	2.66	
t			.785
p			$>.20$
	Traits with 2 indicators		
Mean	13.08	10.50	2.58
SD	2.15	2.21	
t			4.188
p			$<.001$

The data given in Table 1 can only be considered suggestive, since the two groups compared in it were not matched. It does, however, give some indication of the relative efficacy of traits with varying numbers of complex indicators as indices of conflict.

Results

Table 2 shows that no differences in the size of self-concept – ideal-self discrepancies were found between conflictual and nonconflictual traits.

Table 2

Self-Concept–Ideal-Self Discrepancies for Conflictual and Nonconflictual Traits

Statistic	Traits: Nonconflict	Traits: Conflict	Difference
	All subjects ($N = 52$)		
Mean	1.20	1.15	.06
SD	.91	.80	.87
t			.481
p			$>.05$
	Hospital group ($N = 26$)		
Mean	1.30	1.20	.11
SD	.82	.65	.68
t			.782
p			$>.05$
	College group ($N = 26$)		
Mean	1.11	1.10	.01
SD	.39	.48	.54
t			.095
p			$>.05$

A sign test was also done on these data. With the number of cases at 51 (one case was a tie), and the number of the less frequent sign at 21, no significant difference was found between conflictual and nonconflictual traits with respect to discrepancy scores. No significant difference exists between the two groups with respect to the difference between the scores on conflictual and nonconflictual traits.

The results do not support the hypothesis that the presence of conflict over a personality trait is associated with a discrepancy between the concept of self and the concept of the ideal self. Consequently, these results are not in agreement with the findings of Roberts [5] and Bills [1]. However, they are in line with a recent study by Smith [6], who reports that self-concept–ideal-self discrepancies have no discernible effect on ease of perceiving, learning, and remembering of trait adjectives. Perhaps future studies will be able to resolve existing differences.

Summary

This study was designed to test the hypothesis that the presence of conflict over a personality trait is associated with a discrepancy between the concept of self and the concept of the ideal self. A confirmation of this hypothesis would provide a rationale for instruments employing self-concept–ideal-self discrepancies as a measure of adjustment, and of improvement in counseling situations. Fifty-two subjects rated themselves (*a*) as they are, and (*b*) as they would like to be with respect to 25 personality traits on two similar seven-point rating scales. By using the trait adjectives as stimulus words in a word association and reproduction test, six of the traditional complex indicators were obtained, and employed as an index of conflict. The results do not substantiate the hypothesis, and thereby fail to support the contention that discrepancies between the concept of self and the concept of the ideal self are directly indicative of conflict.

Received April 29, 1954.

References

1. Bills, R. E. A validation of changes in scores on the index of adjustment and values as measures of changes in emotionality. *J. consult. Psychol.*, 1953, **17**, 135-138.
2. Bills, R. E., Vance, E. L., & McLean, O. S. An index of adjustment and values. *J. consult. Psychol.*, 1951, **15**, 257-261.
3. Hull, C. L., & Lugoff, L. S. Complex signs in diagnostic free association. *J. exp. Psychol.*, 1921, **4**, 111-136.
4. Raimy, V. C. Self reference in counseling interviews. *J. consult. Psychol.*, 1948, **12**, 153-163.
5. Roberts, G. E. A study of the validity of the index of adjustment and values. *J. consult. Psychol.*, 1952, **16**, 302-304.
6. Smith, G. M. Perceptual defense, learning defense, and memorial defense against threat words. Unpublished doctor's dissertation, Univer. of Rochester, 1954.
7. Thorndike, E. L., & Lorge, J. *The teacher's word book of 30,000 words.* New York: Bureau of Publications, Teachers Coll., Columbia Univer., 1944.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

A Note on Consistency of Rigidity as a Personality Variable[1]

Hermann O. Schmidt, Charles P. Fonda,
and Elizabeth L. Wesley[2]
Norwich State Hospital

Luchins [1] has asserted that problem-solving rigidity is "not a function of the personality per se, but of particular field conditions." On the other hand, findings reported by others suggest that consistent differences in this kind of rigidity do exist among individuals. To test the hypothesis of consistency in the tendency to adhere to induced behavior after it ceases to represent the most direct path to a goal, two different tests purporting to measure this type of rigidity were administered to 93 student nurses. One test, the Wesley Rigidity Scale [3] samples via self-report a variety of behaviors regarded by psychologists as manifestations of rigid behavior. The other test, designed by Luchins [2], involves solutions (to water-jar problems) that give varying evidence of problem-solving (Einstellung) rigidity.

Subjects were classified into three categories of rigidity according to the degree of Einstellung effect: I. *Minimal* — All possible solutions by a short method. II. *Equivocal* — No spontaneous shift to the short method, but consistent use of this method after demonstration. Included in this group also were subjects who either failed to solve one of the "set" problems or solved any of them in an atypical fashion. III. *Maximal* — No spontaneous shift to the short method, with reversion to the longer method at least once after demonstration of the shorter technique.

On the Wesley scale, the mean for 22 subjects in Group I was 19.9 ($SD = 4.89$), whereas for 23 subjects in Group III the mean was 23.35 ($SD = 4.43$). This difference is significant at the .01 level (one-tailed test) with $t = 2.42$. Differences between means for Groups I and II ($t = 1.0$) and between Groups II and III ($t = 1.7$) are insignificant. Thus the hypothesis is supported that rigidity, as here defined, is a consistent personality trait.

Brief Report
Received July 6, 1954.

References

1. Luchins, A. S. Rigidity and ethnocentrism: a critique. *J. Pers.*, 1949, **17**, 449-466.
2. Luchins, A. S. The Einstellung test of rigidity: its relation to concreteness of thinking. *J. consult. Psychol.*, 1951, **15**, 303-310.
3. Wesley, Elizabeth L. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity and of punishment. Unpublished doctor's dissertation, Univer. of Iowa, 1950.

---

[1] An extended report of this study may be obtained without charge from Hermann O. Schmidt, Norwich State Hospital, Norwich, Connecticut, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4396 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[2] Now at the University of Louisville, Medical School.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

A Therapeutic Approach to Reading Difficulties

Julius Seeman
George Peabody College for Teachers
and Benner Edwards
Chicago Public Schools

Reading skill is so central in our culture that much effort has gone into studies of reading in the schools. One aspect of this literature considers the relationship of reading performance to personality organization. From this literature one must conclude that, among the multiple causes of reading difficulty, emotional disturbance should be postulated as a relevant factor in some cases. Many studies which affirm this view are to be found [3, 4, 5, 6, 7, 9, 11].

If we entertain the hypothesis that there is a meaningful relationship between personality disturbance and reading difficulty, it follows as a corollary that experiences which are designed to modify feelings and attitudes in a therapeutic way should also modify reading performance. Bills [1] has shown that when play therapy sessions were held with retarded readers, greater reading gains accrued than during a comparable no-therapy period. In a companion study conducted with well-adjusted retarded readers [2], he found that play therapy made no difference in reading performance. One can conclude that therapy is relevant to reading gain primarily where emotional factors may be at issue, but that therapy will not help performance when emotional factors are ruled out.

The study to be reported here takes these considerations as a point of departure. Specifically, we wish to test the hypothesis that a therapeutic approach to teaching will yield significant changes in personality and in reading performance.

Procedure

Overview of design. In order to test the foregoing hypothesis we must include the following procedures in the experimental design: (*a*) identify children who rank low in personal adjustment and in reading achievement, (*b*) provide an experience which is therapeutic in intent, (*c*) measure the effects of this experience upon personal adjustment and reading performance, and (*d*) provide adequate controls so as to rule out alternative explanations of the experimental outcome. These points will be considered in greater detail below.

Selection of experimental and control groups. The samples used in the study were drawn from the fifth- and sixth-grade classes of a large city. The children were predominantly of lower socioeconomic status.

For the identification of children who ranked low in personal adjustment we used the Tuddenham form of the Reputation Test [10]. This is a peer-rating questionnaire which seeks judgments of specific behavior traits that can be assigned to positive-negative adjustment categories.

For the selection of children low in reading achievement the Gates Reading Survey was used. The selection criterion here was a reading score significantly lower than average expectation for the grade. This criterion is different from the usual criterion of "reading retardation," which is usually defined as reading performance significantly below expectation for a given mental age level. The use of our criterion was predicated on the assumption that where a child was low in both reading and adjustment rank, intellectual functioning might be below potential. This assumption, of course, is implicit in the entire hypothesis of the study. We thus postulate that a therapeutic experience would have a releasing effect upon intellectual functioning, which would be reflected operationally by a gain in reading scores.

Experimental and control groups were selected in two stages: first, the 38 children with the lowest composite rank in both the Tuddenham and the Gates were selected for the study; second, children were paired with reference to score similarity and assigned one each to the experimental and control group. In this way, comparability of experimental and control groups was insured. The first row of Table 1 indicates the close similarity of initial scores.

The pattern of testing. After the groups were selected, the Rogers Personality Test [8] was administered to the children in groups of four to six according to the test's directions. This test was chosen to yield another measure of personal adjustment.

At the end of the experiment, the Rogers and the Gates tests were given again to the experimental and control groups, and the Reputation Test was repeated with the classes from which the experimental and control groups were selected.

The experimental treatment. For the duration of the experiment the teacher-therapist met with the children for one-half hour daily in groups of four to seven. Varied materials were provided, including art materials, books, and games. The teacher made it clear to the children both at the outset and during the experiment that they could use the time as they wished: in talking, reading, working puzzles and games, drawing, or just sitting. The teacher's intent throughout the sessions was to maintain an open, permissive, understanding atmosphere which could encourage exploration and expression by the children.

The children responded by using the time in varied ways. There was frequent interaction between children and teacher. Approximately half the time was spent on books and games, while the other half was divided between the use of art materials and no materials at all. The average number of sessions per child was 67, distributed over a four-month period.

Results

The quantitative findings appear in Table 1. From the table several results are evident. First, it is clear that the experimental group's reading gain is significantly greater than the gain of the control group. The experimental group's gain was .69 years — i.e., about seven-tenths of a year gain in four months.

Table 1

Reading and Personality Test Changes for Experimental and Control Groups

Measure	Gates test		Reputation Test		Rogers test	
	Experimental	Control	Experimental	Control	Experimental	Control
Pretest mean scores	3.80	3.74	28.91	30.64	43.6	43.0
Posttest mean scores	4.49	3.86	16.14	20.23	50.5	43.6
Differences	.69	.12	12.77	10.41	6.9	0.6
p values of between-group differences	$<.01$		.65		$.10>p>.05$	

In the Reputation Test no significant differences are evident. The Rogers test results show a more complex picture. In the Rogers test, the higher the score, the greater the maladjustment. The experimental group shows a pretest vs. posttest difference of 6.9 points in the direction of maladjustment; the control group's change was .6 points. The difference between the experimental and control group changes just misses significance at the 5% level. We therefore conclude that there were no significant differences in Rogers test changes in this experiment, and also take note of the fact that the effect approaches significance in a direction contrary to expectation.

Discussion

The children in our experimental groups showed a relatively strong gain in reading but no gain in adjustment scores. Thus we do not have a simple confirmation or disconfirmation of our hypothesis, but a more complex state of affairs which invites speculative exploration here. Our very first note, of course, is to suggest a repetition of the experiment to determine whether either the reading results or adjustment results can be accounted for by experimental error. We would then know whether we are speculating about fact or illusion. But meanwhile our present results are real enough for the moment, and we shall try to reason out their possible meaning.

A parsimonious conclusion is that a therapeutic experience at school can yield gain in an intellectual function without corresponding gain in measured adjustment. Perhaps the effect of freeing the child to learn can come about as a specific effect without touching the more pervasive personality organization. In other words, perhaps therapy here ameliorated conflict about school learning without a more generalizing effect.

This explanation must remain as one of the possibilities in the situation. On the other hand, we might do well to bring in another facet of the results as a way of seeking other explanations. In this connection we observe that the Rogers test yielded a nearly significant *decrement* in adjustment scores for the experimental group. One way of putting this finding is that there was a greater state of flux in the experimental group results than in the control group results. This fact has a close analogue to the indications in an unpublished manuscript by Bolgar.[1] In a series of studies which analyzed Rorschach protocols before therapy, during therapy, and after therapy, Bolgar found test changes which indicated that, during therapy, disorganization as defined by poor form-quality responses (the "minus" scores) was more in evidence than before therapy. She argued that a loosening-up process might be reflected in these scores. In the present study the greater downward shift in the experimental group scores might be reflecting this same loosening-up tendency—i.e., we may suspect that a tendency to disorganization is part of a therapeutic process. We cannot, of course, do more than speculate about such a possibility, but we should also not like to see this possibility dismissed.

Summary

This study was designed to test the hypothesis that a therapeutic approach to teaching modifies personality and intellectual performance. Comparable experimental and control groups were selected and one group had daily sessions with a teacher-therapist who used therapeutic principles in her work with the group. The experimental group showed significant reading gains as compared with the control group. No significant differences in personality measures occurred. A trend toward decrement in adjustment scores for the experimental group was noted and discussed.

Received March 22, 1954.

References

1. Bills, R. E. Nondirective play therapy with retarded readers. *J. consult. Psychol.*, 1950, **14**, 140-149.
2. Bills, R. E. Play therapy with well-adjusted retarded readers. *J. consult. Psychol.*, 1950, **14**, 246-249.
3. Blanchard, Phyllis. Psychogenic factors in some cases of reading disability. *Amer. J. Orthopsychiat.*, 1935, **5**, 361-371.
4. Ellis, A. Mental hygiene approach to reading disability problems. *J. consult. Psychol.*, 1949, **13**, 56-61.
5. Gann, Edith. *Reading difficulty and personality organization.* New York: King's Crown Press, 1945.
6. Hosey, G. Personality maladjustments resulting from retardation in reading. *Sch. & Community*, 1937, **23**, 273-275.
7. Liss, E. Libidinal fixations as pedagogic determinants. *Amer. J. Orthopsychiat.*, 1935, **5**, 126-131.
8. Rogers, C. R. Measuring personality adjustment in children nine to thirteen years of age. *Teach. Coll. Contrib. Educ.*, 1931, No. 458.
9. Sperry, M. E. A comparison of typical cases and evaluation of treatment of problem children with reading disability. *Smith Coll. Stud. Soc. Work*, 1936, **6**, 289-290.
10. Tuddenham, R. D. Studies in reputation. *Psychol. Monogr.*, 1952, **66**, No. 1 (Whole No. 333).
11. Zirbes, L. Some character and personality problems of remedial cases in reading. *Childh. Educ.*, 1928, **5**, 171-176.

---

[1] Bolgar, Hedda. Personal communication.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

Responses of Normals and Mental Hospital Patients to Freudian Sexual Symbols[1, 2]

Alfred Jacobs
The University of Southern California

The study was concerned with the extent of agreement by normals and mental hospital patients with the gender assigned by Freud [1] to words which purportedly were symbolic of male or female sexual organs.

Subjects consisted of 99 mental hospital patients drawn from five hospital wards, 12 hospital attendants, and 27 student nurses. The *S*s were presented with 113 words which Freud's statements indicated to be typically symbolic of male or female sexual characteristics, and were requested to indicate which words reminded them of the male and which of the female sexual organs.

The agreement with Freud was 65% for admission ward patients, 65% for long-term treatment patients, 48% for postlobotomy patients, 62% for hyperactive patients, and 70% for parole ward patients. The agreement for nurses and for attendants was also 70%.

The percentage agreement with Freud for normals and the total population of patients was significantly greater than chance, but also differed significantly from perfect or 100% agreement. The percentage of agreement with Freud was significantly greater for normals than for postlobotomy patients ($p = .001$), hyperactive patients ($p = .01$), and long-term treatment patients ($p = .001$), and approached significance for admission ward patients ($p = .10$). The superiority of normals to patients occurs on significantly more than one-half the words presented ($p = .01$). The normals showed significantly less blocking than patients in completing the task ($p = .01$).

The study lends support to the conclusion that considerable agreement exists between the gender assigned by Freud to sexual symbols and that assigned by both normals and patients. However, the agreement with Freud, even for groups of *S*s, departs significantly from perfect agreement, and it is possible that cultural factors either in Freud's sample, or in that used in the present study may account for the differences.

The data also lend support to the conclusion that normals agree with Freud to a greater extent than did most of the types of patients studied. It is not clear whether the poorer performance of the patients is attributable to their intellectual impairment, lack of social motivation, or to the presence of more severe or more frequent sexual conflicts.

Brief Report
Received September 14, 1954.

Reference

1. Freud, S The interpretation of dreams. In *Basic writings*. New York: Modern Library, 1938. Pp. 180-549.

---

[1] An extended report of this study may be obtained without charge from Alfred Jacobs, University of Southern California, Los Angeles, California, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4398 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $2.00 for microfilm or $3.75 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[2] Sponsored by the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the author are a result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

Maladaptive Responses to Frustration Predicted from Rorschach Color Responses[1]

Robert W. Harrington
VA Hospital, Marion, Indiana

The proposition that *FC* responses on the Rorschach test indicate a higher level of emotional maturity than do the *CF* and *C* responses has been expressed in various ways. Beck [1, p. 30] has arranged the color responses in an order (*C, CF, FC*) which he considers to correspond to the stages of "emotional maturity" of an individual, and Hertz and Baker state, "The *CF* and *C* together reflect egocentricity and the more primitive, childlike emotions, hence immaturity" [3, p. 13].

Clinical workers have, to a large extent, accepted the above interpretation of color scores and are prone to differentiate, on the basis of Rorschach color scores, between the levels of emotional maturity of otherwise comparable individuals. The purpose of the present study was to test the proposition that *FC* responses are indicative of a higher level of emotional maturity than are *C* and/or *CF* responses.

With the assumption that a positive relationship exists between emotional maturity and adequacy of response to frustration, the hypothesis of the study was: Individuals whose color responses on the Rorschach test are predominantly of the *C* and/or *CF* type will show greater impairment in performance under conditions of frustration than will those individuals whose color responses are predominantly of the *FC* type.

Method

Two groups of *S*s, differing in type of color emphasis on the Rorschach test, were tested under conditions designed to provide for the appearance of habit interference. In this way, since habit interference involves interference with goal-directed behavior, the conditions of frustration were to be fulfilled.

Subjects. The sample consisted of forty, white, male juvenile delinquents at Boys Vocational School, Lansing, Michigan.[2] Two groups of 20 *S*s each were equated on the factors of age and IQ. As can be seen in Table 1, the groups did not differ significantly on these variables.

Table 1

Comparison of Age and IQ Data for Groups with Primary Color Emphasis (*C* + *CF*) and Secondary Color Emphasis (*FC*) on the Rorschach Test
($N = 20$ in each group)

Group	Age		IQ*	
	Mean	Range	Mean	Range
FC	15.85	13.0-17.0	102.75	76-128
C + *CF*	15.75	13.5-18.0	101.35	73-125

* Based on Wechsler-Bellevue scales [4].

In addition, the two groups were equated on several Rorschach variables; namely, $F+\%$, number of *M*, total number of *Y* responses, number of *P*, and total number of responses. All scores were based on individually administered, standard Rorschach examinations.

The independent Rorschach variable was the ratio: Number of *FC* responses/Number of *C*+*CF* responses. As a general rule this ratio had to be at least 2/1 for inclusion in the *FC* group and 1/2 for the *C*+*CF* group. The *FC* group and ratio of Mean Number of *FC*/Mean Number of *C*+*CF* was 3.0/0.35 for the *FC* group and 1.0/3.55 for the *C*+*CF* group, with no subject having a ratio which was different in direction from that of his group. It can be seen from Table 2 that the two groups were roughly equated in regard to the frequency of all the Rorschach factors considered with the exception of the color scores. All Rorschach responses were scored by the investigator according to Beck's system, and the scoring of the color responses was checked by another psychologist.[3] Noncolor responses about which there was some question were similarly checked.

---

[1] This paper is based on a doctoral dissertation submitted to Michigan State College in 1953. The author is indebted to Professors S. Howard Bartley, Alfred G. Dietze, and Albert I. Rabin for their support and criticism throughout the study.

[2] The author wishes to express his appreciation to Mr. Robert Wisner, the superintendent, and to the staff of Boys Vocational School for their cooperation.

Table 2

Comparison of Rorschach Scores for Groups with Primary Color Emphasis ($C + CF$) and Secondary Color Emphasis (FC) on the Rorschach Test

Scores	Group					
	FC		*C + CF*			
	M	*SD*	M	*SD*	diff.	*t*
Total *R*	28.60	8.69	28.55	10.13	.05	.02
Number *P*	7.20	2.29	6.30	1.86	.90	1.36
Number *M*	1.45	1.12	1.70	1.23	.25	.65
Number *Y* (total)	3.20	3.07	2.55	2.48	.65	.72
F + %	81.40	8.82	75.75	11.78	5.65	1.11
Number *C*	.00	.00	1.05	1.43	1.05	3.28*
Number *CF*	.35	.22	2.50	1.28	2.15	7.41*
Number *FC*	3.00	1.30	1.00	1.41	2.00	4.65*

Note.—Corresponding p values were obtained when Fisher's exact test was applied.
* Indicates a p of <.01 (two-tailed test).

Materials and Apparatus

Two experimental tasks were used to compare the performances of the groups under frustration: (*a*) the code-substitution test, and (*b*) the mirror-tracing test. Both tests were administered individually.

The code-substitution test. Two alternate forms of a code-substitution test were used. Both forms utilized the same numbers and geometric figures but the codes, given at the top of the form, were different for Forms A and B. The task of the *S* was to write the appropriate number within each figure as rapidly as possible. The *E* said "wrong" if *S* made a mistake and *S* corrected the error by writing in the correct number over the wrong one. The *S*s were specifically instructed not to erase since to do so would slow them down. Performance was scored for the time per trial and number of wrong substitutions.

The mirror-tracing test. On the mirror-tracing test *S*s had to trace a six-pointed star design under conditions of mirror vision. The design was constructed from copper plates with a chronograph, counter, and stylus connected to the plates to provide a measure of the total time spent off the pathway and the number of times *S* left the pathway. In addition, a stop watch was used to measure the total time spent in tracing the design.

Rationale for the experimental tasks. On both the code-substitution and mirror-tracing tests, it was assumed that previously acquired response tendencies would tend to interfere with the learning of new responses to old stimuli.

On the code-substitution test *S*s were given a number of trials on one form and then tested on an alternate form. Essentially the same conditions existed on the mirror-tracing test, since the previously acquired response tendencies involved in tracing a straight line under direct vision could be expected to interfere with attempts to trace a straight line under mirror vision. Since there was no reason to assume differential amounts of experience among *S*s in making eye-hand movements in space, no training was given in tracing the maze under direct vision.

Procedure. Prior to testing, *E* talked with each *S* and emphasized the facts that *E* had no connection with the institution and that participation in the experiment was entirely voluntary. After ascertaining that *S* was willing to participate, *E* said, "Fine, when we are all through, if I think you have done your best and really tried, I want to give you a candy bar for helping me out" (the candy bars were exposed to the view of *S*), and the reward was again emphasized at the conclusion of the instructions by *E*'s saying, "Remember, if you do your best you get a candy bar." Candy was selected as a reward because of two factors: (*a*) it was acceptable to the institution, and (*b*) candy is highly valued in such a setting.

On the code-substitution test, each *S* was given 18 trials on Form A followed by 12 trials on Form B. Trials were given in groups of three, with two minutes between each set of three trials. During the two minutes between sets *E* engaged *S* in trivial conversation and before each group of three trials *E* said, "Let's see how fast you can do these." After one and one-half minutes of the rest period following trial 18 on Form A had elapsed, *E* placed a copy of Form B in front of *S* and said, "This is different in two ways. First, we will not do as many of these as we did of the others, and secondly, it's different in that the numbers have been changed around, see? Otherwise, it is just like before; if you make a mistake and I see it, I'll say 'wrong.' If I don't see it, but you do, correct it anyway. Remember to work as fast as you can. Ready—go." Except for the number of trials the procedure for Form B was exactly the same as in the preceding 18 trials on Form A.

---

[3] The help of Dr. Durand F. Jacobs is gratefully acknowledged.

Each *S* was then shown the mirror-tracing apparatus and *E* explained its operation (direct vision). Each *S* was then told: "Now you take the stylus and place it at the start (direct vision). When I say 'go' the only place you can look is in the mirror (design was then shielded from direct vision). No matter what happens, you cannot look any place but in the mirror. Ready—go." Many *S*s reached a point where they apparently could not continue tracing and if they gave any indication of discontinuing, *E* said, "Just keep at it, you'll get on to it."

Upon completion of the mirror-tracing task *E* thanked each *S* and gave him a candy bar as the reward which had been mentioned in the instructions. Each *S* was requested not to discuss the tests with any of the other boys and was excused from the room.

Results

Four comparisons were made between groups on the code-substitution test: (*a*) performance on Form A, (*b*) performance on Form B, (*c*) impairment scores (B−A), and (*d*) number of wrong substitutions on each form.

It can be seen from Table 3 that there was no significant difference between the groups on Form A of the code-substitution test. It is also apparent that on the last 10 trials the performance (i.e., habit strength), of the two groups was nearly identical. In view of the results on Form A, the finding that the *FC* group obtained better scores on all 12 trials on Form B is taken to support the hypothesis that the *C*+*CF* group would perform less adequately than the *FC* group under the experimental condition of interference with goal-directed behavior.

Table 3

Mean Scores in Seconds on Each Trial of Two Forms of a Code-Substitution Test for Groups with Primary Color Emphasis (*C* + *CF*) and Secondary Color Emphasis (*FC*) on the Rorschach Test

Trial	Group *FC* Form A	Group *FC* Form B	Group *C* + *CF* Form A	Group *C* + *CF* Form B
1	80.25	94.45	80.50	102.45
2	70.45	79.35	64.85	86.55
3	65.35	77.70	61.25	83.55
4	58.45	75.30	56.30	78.25
5	59.10	75.10	56.45	80.95
6	58.10	72.60	56.00	77.55
7	52.90	69.20	50.35	72.45
8	53.95	68.05	54.60	73.60
9	52.80	66.35	52.50	73.60
10	48.90	60.15	48.75	69.75
11	51.10	61.50	52.25	67.60
12	51.40	60.15	50.80	66.20
13	48.85		49.70	
14	51.35		50.10	
15	51.00		51.10	
16	46.90		48.00	
17	48.40		48.45	
18	49.10		48.75	

Impairment scores were computed by summing the scores for each series of three trials on Form A and subtracting this sum from that obtained for each of the corresponding series of three trials on Form B. A total impairment score, based on all 12 trials, was similarly computed.

Table 4 reveals that on each of the four series of three trials the groups differed significantly, in the expected direction, either in mean scores, variability, or both.

Similarly, a comparison of the groups on total impairment scores reveals significant differences, again in the predicted direction, in both mean scores and variability. These data are presented in Table 5.

When the groups were compared on number of errors (i.e., placing a wrong number in a figure), no significant differences were found.

Table 4

Comparison of Mean Impairment Scores for Each Series of Three Trials on a Code-Substitution Task for Groups with Primary Color Emphasis (*C*+*CF*) and Secondary Color Emphasis (*FC*) on the Rorschach Test

Group	Series (trials)							
	I (1, 2, 3)		II (4, 5, 6)		III (7, 8, 9)		IV (10, 11, 12)	
	M	SD	M	SD	M	SD	M	SD
FC	35.45	22.47	47.35	23.67	43.95	25.65	30.40	18.44
C+CF	65.95	35.68	67.45	41.98	62.20	45.32	51.75	41.88
diff.	30.50		20.10		18.25		21.35	
t†	3.23**		1.86*		1.57		2.09*	
F		2.52		3.15*		3.12*		5.16**

Note.— Where the F ratio was significant the correction indicated by Edwards [2, p. 170] was made in applying the t test.

† One-tailed t test.

* Indicates p of <.05.

** Indicates p of <.01.

On the mirror-tracing test, the two groups[4] were compared as to total time spent tracing the design, number of times the stylus left the pathway, and total time spent off the pathway. None of the differences between the groups approached significance and it must be concluded that the groups could not be reliably differentiated on the basis of their performances on the mirror-tracing test.

Table 5

Comparison of Total Impairment Scores in Seconds on a Code-Substitution Test for Groups with Primary Color Emphasis ($C + CF$) and Secondary Color Emphasis (*FC*) on the Rorschach Test

Group	Mean	*SD*	diff.	*t*†	*F*
FC	157.15	64.12			
			90.20	2.46*	5.24**
$C + CF$	247.35	146.76			

† Corrected for significant F.

* Indicates p of <.02 (one-tailed test).

** Indicates p of <.01.

Thus, while the results of the code-substitution test clearly support the hypothesis, the results of the mirror-tracing test fail to lend such support. Whether the failure of the mirror-tracing task to differentiate the groups can be explained on the basis of the type of behavior involved, the manner in which the test was administered, or whether the reaction to habit interference is itself quite labile remains an open question. The writer is inclined to think, however, that the failure to provide practice under direct vision prior to testing under mirror vision was a serious error in the way the test was administered.

Summary

The purpose of this experiment was to test the proposition that *C* and/or *CF* responses on the Rorschach test are indicative of a lower level of emotional maturity than are the *FC* responses.

Two groups of *S*s, equated on age, IQ, and several Rorschach variables, but differing in type of color emphasis on the Rorschach, were subjected to frustration in the form of habit interference.

On the premise that a positive relationship exists between emotional maturity and adequacy of reaction to frustration, the group whose Rorschach color scores were predominantly of the *C* and/or *CF* type were expected to show greater impairment in performance under frustration than the group whose color responses were predominantly of the *FC* type. Of the two experimental tasks used, only one reliably differentiated the groups. It is concluded that while the results lend some support to the hypothesis, it remains an open question whether the failure of the second task to differentiate the groups is due to the task itself or to the lability of reaction to habit interference.

Received April 6, 1954.

References

1. Beck, S. J. *Rorschach's test.* Vol. II. *A variety of personality pictures.* New York: Grune & Stratton, 1947.
2. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
3. Hertz, Marguerite, & Baker, Elizabeth. Personality patterns in adolescence as portrayed by the Rorschach ink-blot method: II. The color factors. *J. gen. Psychol.*, 1943, 28, 3-61.
4. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

---

[4] One *S* in each group violated the instructions, hence, the *N* for this task is 19 in each group.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

Homosexuality in Paranoid Schizophrenia as Revealed by the Rorschach Test[1]

David Grauer
VA Hospital, Hines, Illinois

The present study was an attempt to test Aronson's [1] finding that the Rorschach records of paranoid psychotics reveal homosexual tendencies. In the course of another investigation described elsewhere [4], the present writer analyzed the Rorschach records of a group of paranoid schizophrenics. Care was taken to eliminate all cases in which there might be reasonable doubt of the adequacy of the diagnosis. This was assured, first, by a study of psychiatric case records to determine the presence of unequivocal evidence of hallucinations and delusions and, second, by the fact that the diagnosis was a "final" one, agreed upon by at least two psychiatrists after a total hospitalization period ranging from three months to over two years.

These criteria differ from those of Aronson, who used ratings of case records. The rating scale was designed to estimate the "extent to which the delusions pervaded the patient's symptomatology" [1, p. 403]. The scale extended from a score of zero, "minimally delusional," to 6, "maximally delusional." Only patients with a rating of at least 4 ("markedly delusional") were selected for the experiment.

The final group of paranoid patients selected by Aronson consisted of 30 individuals, 28 of whom had, in fact, been given a diagnosis of paranoid schizophrenia by the psychiatric staff. Aronson's technique selected out of the total group of paranoid schizophrenics available those whose delusions were considered to have the greatest influence. The published data did not specify the total number of paranoid schizophrenics in the preliminary sample, but the total number of psychiatric records examined was given as 500. If the experience of other psychiatric wards in VA hospitals can be taken as a point of reference, it is estimated that a fairly large proportion of the patients would have been diagnosed as paranoid schizophrenics. The 28 cases finally selected would then constitute a small minority of all the patients thus diagnosed.

The fact that the Freudian hypothesis regarding homosexuality and paranoia could be verified in only a small proportion of cases of paranoid schizophrenia would, in the writer's opinion, seriously limit the applicability of the hypothesis. After all, Freud assumed that latent homosexuality was an important factor in the paranoid group as a whole.

In other respects, the group of 31 paranoid schizophrenics in the present sample closely approximated Aronson's paranoid patients. Both groups were composed of male war veterans, similar in average age (28 years vs. 30.8 years), intelligence (IQ 100 vs. "roughly average intelligence"), and educational level (completion of 10 grades vs. 10.9 grades).

Procedure

The Rorschach protocols of the 31 paranoid schizophrenics in the present investigation were scored for the 21 signs of homosexuality selected by Aronson from the lists of Wheeler [7] and Reitzell [6]. The scoring was done independently by two clinical psychologists including the writer. Differences in scoring between the two psychologists were slight and were easily reconciled after a conference on discrepant points.

In accordance with Aronson's procedure, after the number of homosexual signs present in each patient's protocol had been tabulated, the mean number of signs, mean number of signs based on first response to each card, and mean percentage of signs to total number of responses were calculated. In addition, an attempt was made to determine the prognostic significance of homosexuality, in view of Wheeler's assertion that the finding of homosexuality was generally considered to be an unfavorable sign by psychotherapists. Since the present group of patients was composed of an equal number of improved and unimproved patients, it was possible to ascertain the prognostic value of this factor.

---

[1] From VA Hospital, Hines, Illinois. The writer wishes to express his appreciation to Dr. Roy Brener, Chief Clinical Psychologist, for his assistance in scoring the Rorschach protocols and for his cooperation in making this study possible.

Results

Table 1 shows comparisons between the present data and those for Aronson's normal, nonparanoid psychotic, and paranoid groups. As can be seen, the number and percentage of homosexual signs are definitely lower in the present group than in Aronson's for every comparison. Aronson also reported highly significant differences in homosexual signs between his paranoid cases and his normal group, and between his nonparanoid psychotics and his paranoids ($p = .001$). The differentiation between Aronson's normal patients and the nonparanoid psychotics was much less marked. Most of the latter differences were either not statistically significant or of borderline significance ($p = .02$ to $.10$).

Table 1

Homosexual Signs in Rorschach Records

Group	N	Homosexual signs		Percentage homosexual signs		Homosexual signs on first response to each card	
		Mean	σ	Mean	σ	Mean	σ
Aronson's normals	30	1.10	1.10	4.9	5.4	.57	.92
Aronson's paranoids	30	7.10	4.21	22.9	12.3	3.47	2.05
Aronson's nonparanoid psychotics	30	1.90	1.96	8.5	7.5	.87	.88
Grauer's paranoid schizophrenics	31	3.07	3.14	12.7	11.1	1.36	1.66

The most important comparison is that between the paranoid psychotics and the nonparanoid psychotics. If it can be shown that the paranoids show significantly greater homosexual tendencies than the nonparanoid psychotics, Freud's theory is confirmed. Aronson's finding, previously mentioned, revealed a highly significant difference between the paranoid and nonparanoid psychotics in this respect. The present data do not, however, confirm his results. As can be seen from the table, the mean number of homosexual signs for the paranoid group was 3.07, which differed from the mean number of signs given by Aronson's nonparanoid psychotics by 1.17. The t was 1.75, which failed to reach a minimum p of .05, revealing no statistically significant difference in homosexual signs between the present paranoid schizophrenics and Aronson's nonparanoid psychotics. Similar calculations of percentages and signs on first response on each card revealed t values of 1.71 and 1.44, which were also below acceptable standards of significance.

Another comparison is suggested by Wheeler's original study of Rorschach signs of homosexuality. Wheeler found a median score of between two and three homosexual signs in his group of 100 psychiatric patients in a VA mental hygiene clinic. These patients no doubt included various diagnostic categories with a preponderance of neurotic patients. Of the 31 paranoid schizophrenics in the present study, 17, or 55%, had less than three signs. On this basis, since Wheeler regards total number of signs as the significant factor, there would seem to be considerable overlap in degree of homosexuality shown by the paranoid group and Wheeler's predominantly neurotic patients.

A comparison of the current group with Aronson's group of paranoids shows that of the 31 patients, 28, or 90%, did not attain the mean homosexual score of Aronson's paranoid patients. On the basis of percentages, the ratio of homosexual signs to total number of responses in each record, 25, or about 81%, of the paranoid schizophrenics failed to reach the mean percentage of homosexual signs of Aronson's group of paranoid psychotics.

In view of the commonly expressed opinion that the finding of latent homosexuality constitutes a prognostically unfavorable sign, it was decided to test this hypothesis on the present group, which consisted of improved and unimproved patients. The mean number of homosexual signs per first response to all the cards was 1.31 for the improved group, 1.76 for the unimproved group. The mean percentage of total homosexual responses was 15.6 for the improved group and 15.9 for the unimproved group. It is thus seen that the latent homosexuality, insofar as it can be estimated from the Rorschach record by means of this technique, had no prognostic significance for the group of paranoid schizophrenics.

Discussion

The data failed to confirm Aronson's finding that the Rorschach records of paranoid psychotics exhibit more homosexual content than those of nonparanoid psychotics, as would be expected from Freud's [3] hypothesis regarding paranoia. A comparison of the results with those of Wheeler [7] suggested that, in terms of total number of homosexual signs found, there was little difference in homosexual tendency between the present paranoid schizophrenics and the general run of psychiatric patients in an outpatient mental hygiene clinic. Furthermore, the fact that most of the psychiatric patients (60%) investigated by Wheeler were rated by psychiatrists as having latent or overt homosexual tendencies throws some doubt on the problem. To what extent is the paranoid psychotic unique in showing homosexual trends if most of the members of a heterogeneous group of psychiatric patients exhibit similar tendencies?

According to Fenichel [2], Freud's theory of the relationship between latent homosexuality and paranoia has been confirmed in all psychoanalytic investigations. Yet psychiatrists who have investigated large numbers of paranoid patients challenge this view. On the basis of their experience, Henderson and Gillespie concluded: "The importance of homosexuality in the aetiology of paranoia is not so widespread as the psychoanalytic school would have it" [5, p. 385].

In Aronson's [1] review of the psychiatric literature he reported the incidence of homosexual material found in studies of paranoid psychotics. With one exception, the highest percentage of paranoid patients showing homosexual tendencies was 20. The exception is the report of an analyst who found that 55% of a group of paranoid scizophrenics showed homosexual tendencies in their psychotic behavior. This higher incidence can be explained by the criterion of homosexuality used by the analyst. In addition to the usual behavioral evidence utilized by the other investigators, the analyst included behavior showing "symbolic expressions of homosexuality."

It is thus seen that clear evidence of homosexual behavior is found in only a minority of paranoid patients. If latent homosexuality were of such great importance in the etiology of paranoia as the Freudian hypothesis would indicate, it is difficult to explain why these tendencies do not become overt in the majority of cases of paranoid psychoses. One would expect that, with the breakdown of inhibition and defense that occurs in a psychosis, a larger proportion of paranoids would exhibit frank homosexual material if Freud's hypothesis were correct.

Summary

In this study an attempt was made to test the findings of Aronson that the Rorschach records of paranoid psychotics contain evidence of a higher degree of homosexual tendency than those of nonparanoid psychotics. The subjects in the present investigation consisted of a comparable group of 31 paranoid schizophrenics. The Rorschach records were scored for homosexual content by means of Wheeler's signs. A comparison of the records of the paranoid schizophrenics in the present study with those of Aronson's nonparanoid psychotics reveals no statistically reliable difference. It was also shown that homosexual tendencies, as measured by Wheeler's signs, have no prognostic significance for the group of paranoid schizophrenics. The results of the present study are related to clinical psychiatric findings.

Received April 26, 1954.

References

1. Aronson, M. L. A study of the Freudian theory of paranoia by means of the Rorschach test. *J. proj. Tech.*, 1952, **16**, 397-411.
2. Fenichel, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.
3. Freud, S. Psychoanalytic notes upon an autobiographical account of a case of paranoia (dementia paranoides). In *Collected papers.* Vol. 3. London: Hogarth, 1925. Pp. 387-470.
4. Grauer, D. Prognosis in paranoid schizophrenia on the basis of the Rorschach. *J. consult. Psy-*

chol., 1953, **17**, 199-205.

5. Henderson, D. K., & Gillespie, R. D., *A textbook of psychiatry*, New York: Oxford Univer. Press, 1950.
6. Reitzell, Jeanne M. A comparative study of hysterics, homosexuals and alcoholics using content analysis of Rorschach responses. *Rorschach Res. Exch.*, 1949, **13**, 127-141.
7. Wheeler, W. M. An analysis of Rorschach indices of male homosexuality. *Rorschach Res. Exch.*, 1949, **13**, 97-126.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

Influence of the Preceding Test on the Rorschach Protocol[1]

Robert G. Gibby, Bernard A. Stotsky
VA Mental Hygiene Clinic, Detroit, Michigan
and Daniel R. Miller
University of Michigan

To what extent is the Rorschach protocol influenced by the set created by the preceding administration of another psychological instrument? In clinics, the Rorschach test is typically given as part of a battery. If the protocol should vary with the nature of the previous technique, it would become necessary to standardize the order of presentation.

Recent studies of the Rorschach test point to its sensitiveness to many factors which are not currently controlled in its administration. Protocols differ depending upon the atmosphere in which the test is taken [7], and whether or not anyone has troubled to tell the subject why he is being tested [9]. Also important are variations in the perceptions of the task such as definitions of the purpose of the test and degree of ego involvement [1, 3]. Different examiners, depending upon their sex [4] and needs [8], obtain significantly different scoring categories from comparable groups of subjects [2, 6, 11], and they vary in their interpretations of the same test records depending upon their own personalities [5]. In short, the Rorschach protocol reflects not only the subject's projections but also the definition of the task, the nature of the setting, the examiner's characteristics, and the unique aspects of the relationship between examiner and subject [10].

This paper is concerned with the possible influence on the Rorschach protocol of one aspect of the setting, the experience of just having taken a previous test. The writers were interested particularly in the possibilities that the taking of an intelligence test would create a readiness to give a large number of Rorschach responses, that a thematic instrument might predispose the subject to see motion, that colored blocks would sensitize him to color, and that a drawing test would elicit a set to focus on form.

Procedure

Subjects were selected at random from among patients at the Veterans Administration Mental Hygiene Clinic at Detroit, Michigan. Most of them were neurotics but there were some character disorders and ambulatory psychotics. For purposes of control, it was decided to exclude females, Negroes, organics, disoriented psychotics, and those with IQ's below 80.

The procedure consisted of administering to each *S* one of four initial instruments, the Bender-Gestalt, Thematic Apperception, Wechsler-Bellevue,[2] or Goldstein-Scheerer test, and then the Rorschach test. There were five categories of *S*s with twenty in each category, including a control group that received no test prior to the Rorschach.

Each *S* was referred to one of 16 examiners who was asked to administer the two tests in a particular order and with standardized instructions. The tests preceding the Rorschach were assigned at random before any background information concerning the patients' problems was available. Once the protocols were obtained, they were coded and then scored by one of the writers. Ten records were then rescored by another psychologist for purposes of testing reliability. Eleven variables were tallied: number of responses (*R*), human movement (*M*), pure form (*F*), total shading (*TSH*), total color (*Tot C*), whole (*W*), common detail (*D*), rare detail (*Dd*), human (*H*), animal (*A*), and number of content categories. The average coefficient of reliability of the scoring categories is over .90, and none is below .80.

---

[1] From VA Regional Office, Detroit, Michigan.

[2] In order to eliminate color from this test, the Kohs Blocks were not used during the experiment.

Results

To test the over-all differences among Rorschach protocols administered under these five conditions, analyses were made of the variances of the eleven scoring symbols. None of the results attains or even approaches significance. In fact, almost all of the *F* ratios are close to 1. It can be stated with considerable confidence, then, that for this sample, there is no tendency for an administration of the Bender-Gestalt, Thematic Apperception, Wechsler-Bellevue, or Goldstein-Scheerer tests to influence subsequent performance on the Rorschach test.

Summary

Because of the sensitivity of the Rorschach test to variations in the setting, definition of the task, and examiner's characteristics, it was hypothesized that protocols would vary in accordance with the nature of the initially administered test. No significant differences were found for 11 scoring categories with respect to the five experimental conditions: when the test was given first or when it was preceded by either the Bender-Gestalt, Thematic Apperception, Wechsler-Bellevue, or Goldstein-Sheerer tests.

Received April 19, 1954.

References

1. Calden, G. The relationship of varied test definitions and degrees of ego-involvement to Rorschach test performance. Unpublished doctor's dissertation, Univer. of Michigan, 1951.
2. Cleveland, S. E. The relationship between examiner anxiety and subjects' Rorschach scores. Unpublished doctor's dissertation, Univer. of Michigan, 1950.
3. Cohen, L. The role of two attitudinal factors on group Rorschach performance. Unpublished doctor's dissertation, Univer. of Michigan, 1951.
4. Curtis, H. S., & Wolf, B. The influence of the sex of the examiner on the production of sex responses on the Rorschach. *Amer. Psychologist*, 1951, **6**, 345-346. (Abstract)
5. Filer, R. N. The clinician's personality and his case reports. Unpublished doctor's dissertation, Univer. of Michigan, 1951.
6. Gibby, R. G., Miller, D. R., & Walker, E. L. The examiner's influence on the Rorschach protocol. *J. consult. Psychol.*, 1953, **17**, 425-428.
7. Kimble, G. A. Social influence on Rorschach records. *J. abnorm. soc. Psychol.*, 1945, **40**, 89-93.
8. Lord, Edith. Experimentally induced variations in Rorschach performance. *Psychol. Monogr.*, 1950, **64**, No. 10 (Whole No. 316).
9. Luchins, A. Situational and attitudinal influences on Rorschach responses. *Amer. J. Psychiat.*, 1947, **103**, 780-784.
10. Miller, D. R. Prediction of behavior by means of the Rorschach test. *J. abnorm. soc. Psychol.*, 1953, **48**, 367-375.
11. Sanders, R. The relationship between examiner hostility and subjects' Rorschach scores. Unpublished doctor's dissertation, Univer. of Michigan, 1950.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

An Attempt to Influence the Rorschach Test by Means of a Peripheral Set

Josephine C. Kurtz and Margaret M. Riggs
University of Connecticut

In using a projective technique, the clinician usually starts from the premise that the formally scored variables reflect relatively central or permanent aspects of the personality. He does not expect momentary situational factors or the subject's casual expectations about the task to distort, or even seriously color, these scored variables.

For the Rorschach test, this assumption has come under scrutiny via two main avenues. First, both clinical and experimental evidence [1, 14] indicates that transient but dynamically important personality tendencies may be evoked by the interpersonal relationship induced by the examiner [9], by some strong influence immediately prior to testing [8], or by the test situation itself [7]. In such instances, the Rorschach may reflect the subject's typical reaction in this sort of situation but may not be representative of his usual behavior.

Secondly, there have been several attempts to distort test performance by wholly peripheral, and supposedly non–ego-involving sets. To the extent that this is possible it raises devastating doubts about the premise that the scored variables necessarily reflect central or permanent personality variables. Hutt *et al.* [6] and Gibby [3] told subjects to try to see segments of the blots rather than wholes, and asked them to perceive movement whenever possible. They found significant increases in both perception of details and of movement. Rather than telling subjects directly how to alter their records, Norman, Liverant, and Redlo [11] attempted to induce unconscious sets toward food and human movement by showing a series of picture advertisements and then administering the Rorschach immediately thereafter. These records were compared with Rorschachs given one week earlier or later. The results were completely negative; no increase in food or human movement responses was apparent. Unfortunately, there was no independent proof that viewing the pictures had indeed induced any set. Thus the evidence that the Rorschach is resistant to such temporary sets was suggestive but not conclusive.

In the present study we attempted to set up an unconscious peripheral set to perceive a large number of animals on the Rorschach; the influence was expected to spread to other factors associated with, or to some extent dependent upon, animal percepts. Rorschach himself noted [13] that $A\%$ and $F+$ might be increased, but at the expense of W, M, and O. Lord [9] showed that $A\%$ varied with extent of rapport; Abramson [1] found that $A\%$ changed as a concomitant of a set to see more W or more D. For Hutt [6] $A\%$ was among the less reliable of the variables measured. Thus there was some reason to suppose that $A\%$ might be vulnerable to peripheral influences. We employed the Rees and Israel technique [12] to induce a set to perceive animals. Briefly, subjects are shown pseudo words (e.g., chack) too rapidly for accurate perception. The experimental subjects are told that they would see animals and birds, while the control group is given no special slant. Ordinarily a set is created strongly enough so that without further instruction it carries over into a second task consisting of completing such partial words as S - - l. Thus the second task provides a direct measure of the existence of the unconscious set. The following hypotheses were set up, utilizing this technique:

1. Immediately after demonstrating set on the word-completion task, the experimental subjects will perceive more animals on a group Rorschach test than the controls who received no such set.
2. There will be more P, D, and FM but

fewer *M*, vista, and diffusion responses among the experimental subjects in comparison with the controls.

3. If statistically significant differences are found for others of the conventionally scored variables, these will tend to indicate less creativeness and less anxiety or "inner ferment" among the experimental subjects than among the controls.

If all three hypotheses are proven untenable, clinicians would be supported in the assumption that temporary sets of a peripheral sort do not carry over into the Rorschach test. If the first is supported but not the other two, then such effects as might exist probably would not influence the total personality structure too seriously. If either the second hypothesis (concerned with the direct effect of increased animals on other scores) or the third (embodying the assumption that a set is reassuring to a subject but also dissuades him from trying varied approaches) is verified, then clinicians might be obliged to scrutinize a little more carefully even the fortuitous expectations their subjects bring to a projective task.

Procedure

Three separate laboratory sections of introductory psychology classes were utilized; each was divided randomly into experimental and control groups. In the experimental groups there were 21 males and 10 females; in the control groups there were 24 males and 7 females. The mean age of the control group was 21.5, of the experimental group 22.1, with *SD*'s of 3.32 and 3.38 respectively; the mean difference of .6 was insignificant.

The control groups received the following instructions:

> Eight groups of letters will be shown to you rapidly. The exposure time for each group of letters is very short, so you will have to watch closely. Two practice items will be given to help you adjust yourself to the task. Since the exposure time is very short, set yourself to pay careful attention and to observe the groups of letters as carefully as possible. Do not speak out the answers nor ask any questions until the whole experiment is over. Record below in the proper order what you perceive in each case.

The experimental groups received similar instructions except for the set-inducing phrases:

> Eight words will be shown to you rapidly. The exposure time for each word is very short, so you will have to watch closely. Two practice words will be given to help you adjust yourself to the task. Some of the members of the class belong to the "naive" group and do not know what the words are about. You belong to the "sophisticated" group; hence the following information is given you: Most of the words you are going to see are words having to do with *animals* or *birds*. Set yourself accordingly so that you will perceive as many of the words as possible. Do not speak out the answers, nor ask questions about the words, for it is important that the "naive" group get no hint as to the nature of the words. Record below in the proper order the words which you perceive.

After completing the first task, all groups were introduced to the second task as follows:

> Below you will find 15 skeleton words. Your task is to find real words (not slang nor proper names) which can be made out of the skeleton words by filling in the blanks. In each case you are to record the *first* word that you find fitting the requirements, and you are to see how quickly you can solve the 15 items. You are to use your pencil only in recording each word after it has been discovered; do not write anything during the process of solution. Since this is a speed test, fill in the blanks as quickly as possible. Record your answers and the time taken to complete the list below.

As soon as the second task had been completed, the room was darkened and the group Rorschach was administered, with the following instructions:

> You know when you drop ink on a piece of paper and fold it and open it again, the ink makes weird-shaped blots or splotches on the paper. You will be shown this sort of inkblot projected on the screen. The ones shown have been selected from thousands because they can be seen many different ways by different people. Look at them as they come, and without telling your neighbor, write down on the left-hand side of the paper what they look like, what things you see in them, what they might resemble. You will have three minutes on each slide, to write answers at your own speed. Do you understand?

All questions were answered permissively. After all ten slides had been exposed the inquiry was introduced as follows:

> Now I am going to put each slide back on for about a minute and a half. On the right side of your paper, opposite each response, write down anything more about each response; whatever seemed important to you; just how it was, or what kind it was, or what aspect of the blot was making it look more like that thing than anything else. This is so we can be sure to score what mattered most to you, since different people may notice different things about the very same answer.

After this was done, location sheets were passed out, and subjects were asked to draw a circle around each response if it did not include the whole blot, and label each circle.

This inquiry was modified from the original Harrower-Erickson version [5] in that color, movement, etc. were not suggested, in order to avoid dampening any spontaneous effects of set that might be present.

Without prior knowledge as to which were experimental and which control, the Rorschachs were scored according to the Klopfer system with the addition of Rappaport's combinatory score, and a special "Total Animal" index. This consisted of the number of separate animals mentioned as main, additional, or alternate, expressed as percentage of the number of main responses plus the number of additional animal responses. For all analyses, the three laboratory sections were not differentiated. If fewer than 5 of the 62 subjects scored zero on a particular variable it was considered continuous; for such variables the arc-sine correction for percentages [15] was employed. The resulting distributions were fairly normal; *t* was used as a test of significance only for these variables. For the rest, chi square was used, with 2 by 2 tables constructed according to experimental vs. control and high score vs. low score, the cutting point being that which most nearly divided the total population into halves of equal number [2].

Results

As shown in Table 1, on both Task I and Task II experimental and control groups were quite significantly different; moreover, when Task II is broken into halves, the two groups actually differ more significantly on the second half than the first.[1] Thus a set was not only established in the first task and continued into the second, but the difference between groups was not diminishing.

Table 1

Means, *SD*'s, and Significance of Differences between Experimental and Control Groups for Tasks I and II and for Each Half of Task II

	Mean		*SD*			
	Exper.	Control	Exper.	Control	t	p
Task I (8 items)	5.1	2.6	1.65	1.17	6.77	<.001
Task II (15 items)	9.5	6.3	3.34	2.47	4.22	<.001
Task II, first half (8 items)	5.0	3.5	1.98	1.64	3.20	<.01
Task II, second half (7 items)	4.4	2.9	1.67	1.29	3.89	<.001

In spite of the strength of this demonstration of set in Task II, the results for the Rorschach were essentially negative. As shown in Table 2, no significant differences existed for all those variables amenable to *t*-test analysis. The first seven variables were specified in our hypotheses. Though the *t*'s are of insignificant magnitude, it is interesting that six of the seven are in the predicted direction, the exception being *D%*. An attempt was made to gain more precision by matching subjects for total number of responses, as shown in Table 3, but again all *p* values failed to reach the .05 level of significance.

Table 2

Means, *SD*'s and Significance of Differences between Experimental and Control Groups on Rorschach Variables

Variable	Mean		SD		t	p
	Exp.	Cont.	Exp.	Cont.		
A%	42.4	41.5	7.9	8.8	.417	>.6
Total A%*	50.0	48.4	8.9	10.4	.639	>.5
Animal **P***	3.8	3.4	1.3	1.4	1.128	>.2
Total **P**	4.7	4.4	1.5	1.7	.719	>.4
D%	30.8	31.9	14.1	11.3	- .333	>.7
FM%	22.8	20.9	7.5	9.4	.866	>.3
M%	21.6	24.8	11.1	8.0	-1.279	>.2
F%	38.0	37.2	11.1	10.2	.290	>.7
Combinatory %*	35.3	32.9	11.6	9.9	.864	>.3
W%	52.1	50.7	19.0	9.9	.358	>.7
H%*	27.2	26.7	7.7	7.6	.252	>.8
Object %*	25.3	25.7	10.6	10.0	- .153	>.8
Cards 8, 9, 10%	34.1	33.0	8.0	6.4	.590	>.5
Total no. response	17.3	16.3	6.7	5.6	.626	>.5

*Not part of the ordinary Klopfer scoring system as used here; for total A% and combinatory, see text under "Procedure." Animal **P** is number of **P** seen other than men and bow on Card III, and skin on Card VI. The "object" category is expanded to include all nonliving things, plus nature concepts without plants or animals being specified, e.g., "landscape."

If set produced any effect in the Rorschach, it should be maximal for the upper half of the experimental group on Task II and minimal for the lower half of the control group. Even between these extremes none of the Rorschach differences even remotely approached signifi-

---

[1] The authors are greatly indebted to Dr. Walter Kaess for this and for other statistical and methodological refinements throughout.

Table 3

Significance of Differences on Rorschach Variables with Experimental and Control Groups Matched for Total Number of Responses

Variables specified in hypotheses	*t*	*p*	Variables not specified in hypotheses	*t*	*p*
A%	.214	$>.8$	*F*%	-1.280	$>.2$
Total *A*%	1.334	$>.1$	Combinatory %	.575	$>.5$
Animal *P*	1.067	$>.2$	*W*%	.829	$>.4$
Total *P*	.747	$>.4$	*H*%	.057	$>.9$
D%	$-.969$	$>.3$	Object %	$-.547$	$>.5$
FM%	1.285	$>.2$	Cards 8, 9, 10%	.405	$>.6$
M%	-1.124	$>.2$			

cance. This is an especially devastating refutation of our hypotheses since the low members of the control group not only were given no set toward animals but spontaneously produced relatively few, while the high members of the experimental group may be thought of as having a combination of set to see animals and a natural tendency to do so, one or both being strong.

It was a tenable supposition that if set had produced consistency of behavior, the correlation between number of animals used during the tasks and *A*% on the Rorschach should be higher for the experimental group than for the controls. For Tasks I and II combined versus *A*%, *r* was —.09 for experimental and —.26 for control subjects. For Task II alone versus *A*%, *r* was —.17 for the experimental and —.20 for the control group. These figures are not significant ($p > .05$), but it is curious that the direction should be consistently negative.

Usually we find a modest positive correlation between *F*% and *A*%. If this relationship were destroyed among the experimental group but preserved among the controls, it would suggest that set had had a distorting effect. The *r* between *F*% and *A*% for the experimentals was +.03, and for the controls +.32; again neither is significant, though as in the bulk of the data the trend is in the predicted direction.

When all of the Rorschach material was subjected to chi-square analysis, i.e., those variables considered continuous and those not, of 56 analyses, six were significant ($p < .05$) when fewer would have been expected by chance alone. These six are presented in Table 4.

Table 4

Chi Square and *p* Levels for Those Rorschach Variables Yielding Significant Differences between Groups

Rorschach variables	χ^2	*p*	Direction for exp. group
Total number of animals	4.15	$<.05$*	High
Add. animal responses	4.35	$<.05$	Absent
$A + H$ in the same response	5.02	$<.05$	Absent
Rare detail (*dr*)	5.02	$<.05$	Absent
Formed color (*FC*)	4.01	$<.05$	Absent
Unformed color ($CF + C$)	6.80	$<.01$	Present

*Of the variables for which predictions were made only the total number of animals yielded a significant relationship; because of the one-tailed nature of the hypotheses, p may be considered >.025 for this variable.

Although the *t* for total number of animals mentioned was not significant, more experimental than control subjects were above the median ($p < .05$). This is the only straightforward indication of the effect of set in all the material. It must be remembered, however, that total number of animals mentioned as defined in this study, is not a traditional Rorschach variable. Unexpectedly, it was the control subjects who were apt to use additional responses containing animals, and who used animals fused with humans in the same response. Thus if the experimental subjects used animals at all, they were likely to make them the main content. It is not obvious why it should have been the controls who used the rare locations (*dr*), unless indeed set had the narrowing effect predicted in Hypothesis 3. There seems to be no ready explanation for the prevalence of amorphous color among the experimental group and formed color among the control group; it is tempting to attribute these last two to chance!

Discussion

The evidence seems to indicate that percentage of animals seen on the Rorschach blots is not vulnerable to distortion by a temporary set, though there were a priori reasons for suspecting it to be inherently less stable than many other Rorschach variables. Unfortunately, this experimental design did not entirely rule out the possibility that the set, having carried over strongly to a second task, simply vanished in the face of a third task. All that we can say is that it was certainly present immediately prior to the Rorschach and not clearly demonstrable during it. If a fourth task had been presented after the Rorschach we do not know whether experimental and control groups would have been differentiated again. It seems to be the nature of sets, however, to persevere and increase in strength the longer they prove successful for problem solving [10]. We find it difficult to believe that the set to see animals simply dissipated itself spontaneously in the minutes between Task II and the Rorschach. All the evidence suggests, therefore, that the Rorschach evoked from all the subjects their characteristic behavior so strongly as to override the existing set; in effect, the subject followed some more fundamental personal tendency, quite as Rorschach workers assume. The very fact that the insignificant data showed a persistent trend in the predicted direction argues against the supposition that no set existed at the moment when the Rorschach was presented, while the failure of the effect to reach any great magnitude testifies to the genuine coerciveness of the Rorschach material. So far as this study is concerned, clinicians remain secure in their assumption that implicit peripheral sets will not influence Rorschach results to any appreciable degree.

Summary and Conclusions

1. College students were randomly divided into a control group and an experimental group, the latter being given a set to perceive animals by the Rees and Israel technique [12]. On an initial task (word recognition) the two groups were significantly differentiated and remained so on a second task (word completion) with $p < .001$. Immediately thereafter all subjects were given the group Rorschach.

2. It was our hypothesis that having taken the set, the experimental subjects would perceive more animals on the group Rorschach test than the controls.

3. Despite clear indications of set in Task II, the Rorschach results were essentially negative. No significant differences were found for any of the variables for which t test was appropriate; when cases were matched for total number of responses the p levels were no higher.

4. Initially we had predicted that among the experimental group there would be higher $A\%$, more animals mentioned, higher $D\%$, more populars, more FM, but fewer M, vista, and diffusion responses. Although, except for $D\%$, the obtained differences were in the predicted direction, none reached statistical significance.

5. When all scored variables were treated by chi square, of the 56 analyses only six achieved significance, and only two of these even remotely supported any of the original hypotheses.

6. All our evidence suggests that the Rorschach material was genuinely coercive, evoking from the subjects their characteristic behavior and overriding a strongly established pre-existent set. So far as this study is concerned, Rorschach workers remain secure in the assumption that implicit peripheral sets will not influence test results to any appreciable extent.

Received August 25, 1954.
Early publication.

References

1. Abramson, L. S. The influence of set for area on Rorschach test results. *J. consult. Psychol.*, 1951, **15**, 337-342.
2. Cronbach, L. J. Statistical methods applied to Rorschach scores. *Psychol. Bull.*, 1949, **46**, 393-429.
3. Gibby, R. G. The stability of certain Rorschach variables under conditions of experimentally induced sets: 1. The intellectual variables. *J. proj. Tech.*, 1951, **15**, 3-26.
4. Gibby, R. G., Miller, D. R., & Walker, E. L. The examiner's influence on the Rorschach protocol. *J. consult. Psychol.*, 1953, **17**, 425-428.

5. Harrower-Erickson, Molly R. Directions for administration of the Rorschach group test. *Rorschach Res. Exch.*, 1941, **5**, 145-153.

6. Hutt, M. L., Gibby, R. G., Milton, E. O., & Pottharst, K. The effect of varied experimental "sets" upon Rorschach test performance. *Rorschach Res. Exch.*, 1950, **14**, 181-187.

7. Kimble, G. A. Social influence on Rorschach records. *J. abnorm. soc. Psychol.*, 1945, **40**, 89-93.

8. Levine, K., Grassi, J. R., & Gerson, H. Hypnotically induced mood changes on the verbal and graphic Rorschach. *Rorschach Res. Exch.*, 1943, **7**, 130-144.

9. Lord, Edith, Experimentally induced variations in Rorschach performance. *Psychol. Monogr.*, 1950, **64**, No. 10 (Whole No. 316).

10. Luchins, A. S. Mechanization in problem solving: the effect of Einstellung. *Psychol. Monogr.* 1942, **54**, No. 6 (Whole No. 248).

11. Norman, R. D., Liverant, S., & Redlo, Miriam. The influence of a superficial immediately preceding "set" upon responses to the Rorschach. *J. consult. Psychol.*, 1952, **16**, 261-264.

12. Rees, Hulda, J., & Israel, H. E. An investigation of the establishment and operation of mental sets. *Psychol. Monogr.*, 1935, **46**, No. 6, (Whole No. 210), 1-26.

13. Rorschach, H. *Psychodiagnostics.* (3rd Ed.) Berne, Switzerland: Huber, 1942.

14. Schachtel, E. G. Subjective definitions of the Rorschach test situation and their effect on test performance. *J. Psychiat.*, 1945, **8**, 419-448.

15. Snedecor, G. W. *Statistical methods.* (3rd Ed.) Ames, Ia.: Iowa Collegiate Press, 1940.

Journal of Consulting Psychology
Vol. 18, No. 6, 1954

Inter-Judge Agreement on Traits Rated from the Rorschach[1]

Leonard Gelfand, Bruce Quarrington, Harley Wideman, and Jean Brown
Department of Psychiatry, University of Toronto

One concomitant of the recent trend toward global validation of projective techniques is an attempt to utilize clinicians' judgments in the design of validation studies. Hence it is necessary to establish the extent of agreement between judges. This paper reports a study to determine the correlation between the ratings of six clinicians who rated 36 patients on 10 traits from the Rorschach.

All six raters were skilled clinical psychologists, had received recognized Rorschach training and supervision, and had from four to eight years experience with the technique. They participated in the research design and hence knew the purpose for which they were rating the subjects.

The 36 patients were selected by the raters from the closed case file at the Toronto Psychiatric Hospital. One-half were inpatients and one-half outpatients. The age range was from 17 to 62 with a mean of 30.03 and a median of 27.5 years. The subjects fell into the following diagnostic categories: anxiety state, 8; depression, 3; obsessive-compulsive, 1; undifferentiated psychoneuroses, 7; adolescent behavior disorder, 2; undifferentiated schizophrenia, 6; preschizophrenic disorders, 4; paranoid schizophrenia, 1; affective psychoses (manic-depressive, psychotic depression, involutional disorders), 4.

The 10 traits rated were: desire to gain insight, feeling of physical inadequacy, dissatisfaction with self, impulsivity, dissatisfaction with life situation, basic intellectual potential, assertiveness, present intellectual functioning level, identification with social role of own sex, and use of wish-fulfilment fantasy. Each trait was rated on a 10-point scale.

Patients were rated individually by each judge. The information available to the judge was the complete, scored Rorschach, the age and sex of the patient, whether he was a hospital or clinic patient, and which of their colleagues administered and scored the Rorschach.

Product-moment correlations were calculated for the ratings of each judge with every other judge on each trait. The 150 intercorrelations ranged from —0.116 to +0.716 with the estimated median about 0.30 and nine negative correlations.

The correlations are judged as too low to indicate consistency or agreement between the raters' judgments and hence no tests of significance have been applied. It seems reasonable to conclude that the judges disagree in their ratings of the patients. The disagreement may be the result of three factors: (*a*) the heterogeneity of the patient sample, (*b*) the complexity of the material judged, and (*c*) the lack of adequate definition of trait terms coupled with the use of differential Rorschach criteria for judging these traits.

Brief Report
Received September 13, 1954.

---

[1] An extended report of this study, which was financed by a Provincial-Federal Health Grant, may be obtained without charge from L. Gelfand, 241 Elizabeth St., Toronto, Ontario, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4397 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

NEW BOOKS AND TESTS

Books

Anastasi, Anne. *Psychological testing.* New York: Macmillan, 1954. Pp. xiii + 682. $6.75.

The field of testing has grown so rapidly that most contemporary books on psychological testing are either extensive and superficial, or intensive and incomplete. Anastasi's new college text achieves a nice compromise between these extremes. This is not a compendium of present-day tests, each briefly described and briefly criticised. Neither is it a recapitulation of the manuals of a few widely used tests. It is a carefully planned and integrated volume, which begins with a consideration of the principles of psychological testing, proceeds to employ carefully selected examples of tests of general classification, aptitude, and achievement, and ends with measures of personality characteristics ranging from inventories through projective techniques and situational tests. In every instance, the instruments selected for illustration are representative, and their advantages and shortcomings are handled critically. Although the book is primarily a college text, it cannot fail to attract the attention of practicing psychometricians and clinicians. Particularly important for both students and practitioners are the sections on the ethics of control of tests, on test validity, on the limitations of infant and preschool tests, and on the peculiar characteristics of projective techniques and situational tests. — *A.M.G.*

Beck, Samuel J. *The six schizophrenias, reaction patterns in children and adults.* Research Monograph No. 6. New York: American Orthopsychiatric Association, 1954. Pp. xi + 238. $5.00.

This research monograph reports an interdisciplinary study of outpatient schizophrenics, based on systematic clinical descriptions by psychiatrists and Rorschach interpretations by psychologists. The major conclusions come from three inverse factor analyses of *Q* sorts of 120 descriptive items. One factor analysis was of 12 patients, 7 adults and 5 children, on whom the psychiatrists and psychologists agreed; a second was of the 8 patients, 5 adults and 3 children, on whom they disagreed. The third inverted factor analysis was of 20 children, "some" (number unreported) of whom were included in the two preceding analyses. The first analysis revealed three types of schizophrenia, circumspectly labeled S-1, S-2, and S-3, which might be characterized briefly as intellectually disrupted, affectively disturbed, and rigidly controlled. The second analysis, of the eight "disagreement" cases, disclosed two more types, SR-1 and SR-2, which were differentiated only by Rorschach patterns, not by clinical descriptions. The analysis of the children revealed one further type, S-C. Studies are also reported of the comparison of the Rorschach scores of schizophrenics, neurotics, and normal controls, of the development of schizophrenia in children based on Rorschachs 2 to 12 years apart, and of environmental conditions associated with schizophrenia. Although the research uses refreshingly original approaches and represents an impressive amount of labor, it falls short of achieving its potentialities. First, insufficient use is made of the normal control data. One would like, for example, to know how many Rorschachs of clinically normal persons would fall into one of the "schizophrenic types." Second, the data are not always optimally reported or analyzed. As one instance, the criteria for choosing the crucial "agreement" and "disagreement" groups of patients are not given, and the reviewer's digging into correlation matrices in an appendix unearthed the distressing observation that the psychiatrists and psychologists seemed in better agreement on 3 of the 8 "disagreement" cases than on 3 of the 12 "agreement" ones. Reliability is treated with more optimism than data; again, the appended tables reveal that the correlations between the *Q* sorts made by two Rorschach interpreters for 15 cases ranged from .01 to .65 with a median of .50, certainly not an impressive scoring reliability. Third the need for cross validation is not recognized. Fourth, in spite of the author's pleas for idiographic research, there is question of the generality of findings based mainly on three inverse factor analyses of 12, 8, and 20 cases which included only 12 adults, all outpatients. The study makes con-

---

Note.—The reviews were prepared by the Editor and Associate Editors, who may be identified by their initials.

tributions to method and proposes useful hypotheses for further research. It almost surely does not, as the title implies, identify *the* six schizophrenias. — *L.F.S.*

Burt, Cyril. *The causes and treatment of backwardness.* New York: Philosophical Library, 1953. Pp. 128. $3.75.

The author describes this book as an abridgment and revision of his earlier *The Backward Child* of fifteen years ago. The new volume is primarily concerned with educationally substandard children. We think it regrettable that the "educationally" was not incorporated in the new title, since many readers will anticipate that the book deals with the mentally backward. Actually there is more need for exposition of the multiple causes of scholastic retardation than of mental subnormality, and the more specific title would have drawn a wider audience and supplied a more useful purpose. We also wish that the author might have elaborated the meaning of "educational" backwardness, since slow learning has directions as well as degrees. We regret this more in the light of the author's authority, insight, and experience which give this exposition both breadth and merit, and which in spite of beguiling simplicity, impresses us as excellent comprehension of children's learning barriers. — *E.A.D.*

Cronbach, Lee J. *Educational psychology.* New York: Harcourt, Brace, 1954. Pp. xxvii + 628. $7.50.

Cronbach has made an excellent integration of the resources of psychology with the needs of education. Psychologists will like this book because of its sound but simple handling of evidence and theory; educators will like it because of the clear awareness of the realities of the classroom. Students will enjoy its lucid writing and attractive format. — *L.F.S.*

English, O. Spurgeon, & Finch, Stuart M. *Introduction to psychiatry.* New York: Norton, 1954. Pp. viii + 621. $7.00.

Intended as a beginning text for medical students, *Introduction to Psychiatry* can serve as well to orient graduate students in psychology and related social sciences to modern psychiatry. Avowedly "dynamic" and Freudian, the treatment in most of the volume is essentially eclectic. It follows rather closely the Classification of Mental and Emotional Diseases adopted by the American Psychiatric Association in 1952. In addition to the systematic treatment of personality disorders, there are excellent, succinct chapters on personality development and structure, child psychiatry, principles of psychotherapy, and mental hygiene. The chapter on mental deficiency is weak. A tremendous amount of psychiatric material is covered in this text, most of it necessarily brief — perhaps too brief for reference purposes. It was designed, however, as an "introduction," and as such is excellent. — *M.K.*

Hodgson, Kenneth. *The deaf and their problems.* New York: Philosophical Library, 1954. Pp. xx + 348. $6.00.

The author has divided the book into three sections. The first part is on the ear and the mechanism of hearing. Part two is a historical survey of the problems of the deaf from ancient and medieval times. Part three is a discussion of the problems of the deaf in the twentieth century. Because we often learn from the mistakes of the past, the book furnishes excellent background material for one who is interested in the problems of the deaf and in modern-day solutions. The author strongly favors the view that the conflict between the oral and silent methods of teaching the deaf should have been allowed to die out; there should be experimental effort on an international scale on the correct teaching of the deaf. While work in America is mentioned, the author seems much better informed on the methods used in Europe and the British Isles. The book is strongly oriented toward English problems and schools. — *B.M.L.*

Hurlock, Elizabeth B. *Developmental psychology.* New York: McGraw-Hill, 1953. Pp. ix + 556. $6.00.

"The purpose of this book is to give as complete a picture of the developmental changes of the total life span of the human being as is possible within the two covers of one book." These changes are primarily psychological, with only sketchy reference to the medical and social aspects. This volume apparently telescopes material from the author's earlier successful books on child and adolescent development, with the addition of chapters on adulthood and old age. The contents are impressive for both generalization and detail. The coverage is broad and eclectic yet skillfully discriminative. Obviously no serious depth could be included throughout, but one gets the impression of an adequate and scholarly survey of a field which has rapidly reached its own fruitful maturity. — *E.A.D.*

Lazarsfeld, Paul F. (Ed.) *Mathematical thinking in the social sciences.* Glencoe, Ill.: Free Press, 1954. Pp. 444. $10.00.

These papers, by eight authors, were selected as examples of situations which might typically develop between the social sciences and mathematics. This is certainly an important book, but most of it is extremely difficult for the social scientist with no more than the "modicum of training in mathematics" suggested by the editor as sufficient. Anderson develops a model for analyzing changes in attitudes over time on the basis of analysis of stochastic processes. Its application is demonstrated with a panel study of potential voters who were interviewed six times during an election campaign. More general models are also developed. Rashevsky expresses basic problems in interpersonal contacts in terms of differential equations. His first paper deals with imitative behavior and the second with the distribution of status and its alteration through repeated contacts. Coleman gives a further explication of Rashevsky's models, including two others in his analysis, and shows specific relations between the mathematical elements of the models and the elements of the concrete social situations. Marschak discusses the role of probability theory in the social sciences with emphasis on "subjective probability." As an example of the relation between probabilistic models and social problems he analyzes the effect of a government tax policy. Guttman's first contribution is a development of the psychological implications of the mathematics of his scalogram analysis, already well known to psychologists. The second presents a generalized form of factor analysis. Lazarsfeld's paper concerns the relations between social research procedures and the basic equations of latent structure analysis. Simon contributes a general discussion of model building in the social sciences. — *A.R.*

Lorand, Sandor. (Ed.) *The yearbook of psychoanalysis.* New York: International Universities Press, 1954. Pp. 350. $7.50.

The ninth volume of the Yearbook resembles its predecessors in furnishing a range of papers, reprinted from other sources, representing both theoretical and clinical contributions to psychoanalysis. While many of the writings expand, interpret, or paraphrase Freud, the total sweep of the volume is such that a diversity of contemporary viewpoints is provided. In the reviewer's opinion, the papers by the Bernfelds ("Freud's First Year in Practice"), by Klein ("The Origins of Transference"), by Szasz ("On the Psychoanalytic Theory of Instincts"), by Lewin ("Phobic Symptoms and Dream Interpretation"), by Brewster ("Separation Reaction in Psychosomatic Disease and Neurosis"), by Rangell ("The Analysis of a Doll Phobia"), and by Roheim ("The Evil Eye") illustrate most clearly the current developments and controversies selected for inclusion. The volume makes available in convenient form a great many contributions to psychoanalysis which otherwise might well go unnoticed by those not directly in the analytic field.—*A.M.G.*

Pickford, R. W. *The analysis of an obsessional.* New York: Norton, 1954. Pp. xii + 234. $4.00.

Valuable because of the extensive material it presents on a single case, this book is a disappointment on three scores. First, the record of the analytic interviews is a reconstruction from the therapist's notes, not an electrical transcription. Consequently, in spite of the richness of the reported observations, readers must wonder about the extent to which they are colored by the analyst's preconceptions and theoretical orientation. Pickford's remarkable statement that "The real remedy for the supposed weakness in this lack of 'objective' recording is surely to be psycho-analysed oneself," made in defense of his not taking a verbatim record, represents the second source of disappointment. The book has the tone of a defensive dogmatism that discourages critical thought and tends to stereotype the insightfulness that case presentations can often elicit. Third, although it is written by a distinguished psychologist with a brilliant record in experimental fields, *The Analysis of an Obsessional* shows no sign of its author's professional identity. The potential contributions of psychology appear to have been thoroughly forgotten in favor of a wholesale capitulation to psychoanalytic orthodoxies. While this volume may have heuristic importance as an illustration of one way of dealing with a particular kind of psychotherapeutic problem, it seems quite lacking in novelty, inventiveness, and the refreshing tough-mindedness that one can legitimately demand from this kind of literature, especially when prepared by an outstanding psychologist. — *E.J.S.*

Sarason, Seymour B. *The clinical interaction, with special reference to the Rorschach.* New York: Harper, 1954. Pp. x + 425. $5.00.

Sarason has produced a unique volume and a most stimulating one. The book is intended to serve as a text for instruction in the Rorschach and was developed from the author's classes. But it is much more than a Rorschach book. Its thesis is that the Rorschach cannot be understood in isolation, but only in terms of the interpersonal interactions of the clinical situation. The first eight chapters, therefore, do not deal with the test at all, but with the variables inherent in interpersonal relations — the stimulus task, the time and place, the clinician as a person, his age and sex, and the attitudes and expectations of the client. In discussing these variables, Sarason draws upon experimental studies as well as case illustrations from social psychology, psychotherapy, and psychodiagnostics. Only on page 109 does the student meet the Rorschach itself, in a well-written series of chapters on instructions, location, color, shading, movement, form, content, and integrated interpretation. Extensive use is made of Rorschach research findings; no other work on the Rorschach supports its doubts and claims with so much data. The negative nature of much of the evidence is faced frankly. The text concludes with an analysis of six cases, designed to illustrate the processes of inference by which the psychologist draws his conclusions. The volume quite evidently leaves its author conflicted — a feeling that will be shared by many readers. When we remove from the Rorschach all of the beliefs controverted by evidence, the remaining substance is thin. Instructors who like their students to think, instead of to be indoctrinated in a ritual, may well consider this book seriously. — *L. F. S.*

Stolz, Lois Meek. *Father relations of war-born children.* Stanford, Calif.: Stanford Univer. Press. 1954, Pp. viii + 365. $4.00 (Paper).

This study of the adjustments of fathers and their first-born children to stress determined by the return of the father from overseas is outstanding for its testing of clinically conceived hypotheses by means of objective and appropriate methods. Nineteen families, separated by war during the first pregnancy of the mother and reunited after the first child was at least one year old, constitute the major group for study. These families are compared with control groups of families in which separation did not occur. The major hypothesis of the study is that the war-separated families show adjustment difficulties, particularly between father and child, and that the father's attitude affects the development of the child adversely. Methods of testing this and other hypotheses were devised with care, and included interviews with fathers and mothers, free observations of children with peers and adults, and studies of children in projective play situations, such as aggression with balloons and blocking, doll play, story completions, and dramatic play completions. The major and minor results of the study are numerous and highly significant for our understanding of personality development in early infant-parent interaction. The dynamic pattern of father-child rivalry for the mother, withdrawal by the child, and consequent paternal distress shown in the development of authoritarian attitudes, for example, provides an important contribution to our knowledge of how adjustive techniques evolve. — *A. M. G.*

Sullivan, Harry Stack. *The psychiatric interview.* New York: Norton, 1954. Pp. xxiii + 246. $4.50.

In this volume Otto Will has organized some of the lectures and writings on interviewing by the late Harry Stack Sullivan. Dr. Will has done an excellent editorial job and contributed a helpful introduction. Sullivan's comments are both learned and lively. He thoroughly understood the interpersonal dynamics of interviewing, and writes of it with evident enjoyment. His style is penetrating and salty, combining a feeling for the theoretical significance with a flare for earthy interpretation. The clinician who reads this book will learn a lot about interviewing and a lot about Harry Stack Sullivan, and he will have a delightful time doing it. — *W. A. H.*

Swartz, Harry. *The allergic child.* New York: Coward-McCann, 1954. Pp. 297. $3.95.

Most of us are perfunctorily alerted to the discomforts and annoyances of allergic idiosyncracies. But relatively few laymen appreciate the multiple consequences, both direct and secondary, associated with such morbid sensitivities. Hence this book may well impress some readers as rabid and even propagandistic. There is, indeed, an initial aura of patent-medicine advertising that creates fear or at least skeptical apprehension which then changes to earnest concern. The widespread ramifications of body-system damage are perhaps less understood than the dynamic, behavioral, and learning sequelae. The reviewer is unequal to appraising the strictly somatic significances, but anyone touched with allergy or associated with allergy-prone children will acquire a more informed and determined view of this pervasive distress. The child-guidance therapist will find herein one more and highly significant desideratum in his appraisals of the problems of children and adults. — *E. A. D.*

Westwood, Gordon. *Society and the homosexual.* New York: Dutton, 1953. Pp. 191. $3.00.

"This is not a medical treatise, but an attempt to evaluate the social implications of homosexuality". It "includes sexual activities . . . between two adult men. It does *not* include infantile and early pubertal homosexuality . . . and it does *not* include the pseudo-homosexuals or the infanto-homosexuals." Its "case," as noted by Dr. Edward Glover's introduction, is "in brief, that homosexuality . . . is not . . . the lecherous perversion of self-indulgent degenerates, but . . . a powerful unconscious force to which, in other forms, civilization owes much of its strength and some of the greatest of its achievements." In short it offers a sympathetic exposition of the more recently tolerant and analytically insightful interpretations without, it seems to the reviewer, adding materially thereto. — *E. A. D.*

Test

Andrew, Gwen, Hartwell, Samuel W., Hutt, Max L., & Walton, Ralph E. *Michigan Picture Test.* Ages 8-14. Individual or small-group test. 1 form. Untimed. Set of 16 plates ($6.00); analysis sheets ($1.60 per 20); Manual, pp. xxiii + 108, ($2.00). Also, *Rating Scale for Pupil Adjustment* ($1.15 per 20); with manual, pp. 4; specimen set ($9.00). Chicago: Science Research Associates, 1953.

The *Michigan Picture Test* is a new apperception test for children 8 to 14 years of age, developed through seven years of research by the Michigan Department of Mental Health. No other picture-story test has been accompanied by so much data relevant to its standardization and validation. The test may be used in several ways. For screening, four "core" pictures are used, which represent a family situation, a peer-group situation, lightning at night, and a blank card. Responses to these four plates may be scored for variables validated against a criterion of teacher ratings of adjustment — tension index, verb tense, and direction of forces — from which a combined maladjustment index may be obtained. Scoring reliability is good, but there are no data on subject response reliability. The analysis sheet provides space for recording four more variables quantitatively — interpersonal relations, personal pronouns, popular objects, and psychosexual level — but users are warned that these scores bore no significant relation to the criterion. An entire series of twelve plates, four of which have separate cards for boys and girls, may be used for full clinical study. The well-prepared manual contains an essay by Hutt on the psychology of picture tests, and a full account by the other authors of the selection of the pictures and the validation and standardization of the test. The Michigan pictures provide a useful clinical tool for an age group hitherto ignored in the construction of apperception tests.—*L. F. S.*

Books Received

Arbuckle, Dugald S. *Student personnel services in higher education.* New York: McGraw-Hill, 1953. Pp. x + 352.

Argyris, Chris. *Organization of a bank.* Studies in Organizational Behavior, No. 1. New Haven, Conn.: Labor and Management Center, Yale Univer., 1954. Pp. vii + 282 (paper).

Arsenian, Seth, & McKenzie, Francis W. *Counseling in the YMCA.* New York: Association Press, 1954. Pp. 126. $2.00.

Crow, Lester D., & Crow, Alice. *Readings in general psychology.* New York: Barnes & Noble, 1954. Pp. x + 437 (paper). $1.75.

Daniels, Edgar E. *Dark destiny.* New York: Vantage Press, 1954. Pp. 172. $3.00.

Driver, Helen I. *Multiple counseling, a small-group discussion method for personal growth.* Madison, Wis. (803 Maygara Rd.): Monona Publications, 1954. Pp. 280. $5.00.

Eidelberg, Ludwig. *An outline of a comparative pathology of the neuroses.* New York: International Universities Press, 1954. Pp. 263. $4.50.

Fink, Harold K. *Long journey.* New York: Julian Press, 1954. Pp. xviii + 298. $3.95.

Gruenberg, Sidonie M. (Ed.) *An encyclopedia of child care and guidance.* Garden City, N. Y.: Doubleday, 1954. Pp. 1016. $7.50.

Hayes, E. Nelson. (Ed.) *Directory for exceptional children.* Boston (11 Beacon St.): Porter Sargent Publishers, 1954. Pp. 96 (paper). $1.10.

Hilden, Arnold H. *Manual for Q-sort and random sets of personal concepts.* Webster Groves 19, Mo. (628 Clark Ave.): Author, 1954. Pp. 35 (paper). $2.00.

Kendall, Patricia. *Conflict and mood, factors affecting stability of response.* Glencoe, Ill.: Free Press, 1954. Pp. 182. $3.50.

King, H. E. *Psychomotor aspects of mental disease.* Cambridge: Harvard Univer. Press, 1954. Pp. xiv + 185. $3.50.

Knight, Robert P. (Ed.) *Psychoanalytic psychiatry and psychology.* Clinical and Theoretical Papers, Austen Riggs Center, Vol. 1. New York: International Universities Press, 1954. Pp. vii + 391. $6.00.

Leuba, Clarence. *The sexual nature of man, and its management.* Doubleday Papers in Psychology, No. 6. New York: Doubleday, 1954. Pp. vii + 40 (paper). 85¢.

McAuley, W. F. *The concept of schizophrenia.* New York: Philosophical Library, 1954. Pp. 145. $3.75.

Maslow, A. H. *Motivation and personality.* New York: Harper, 1954. Pp. xiv + 411. $4.50.

Meenes, Max. *Studying and learning.* Doubleday Papers in Psychology No. 9. New York: Doubleday, 1954. Pp. viii + 68 (paper).

Moses, Paul J. *The voice of neurosis.* New York: Grune & Stratton, 1954. Pp. v + 131. $4.00.

Rogers, Carl R., & Dymond, Rosalind F. (Eds.) *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. x + 447. $6.00.

Sanderson, Herbert. *Basic concepts in vocational guidance.* New York: McGraw-Hill, 1954. Pp. xiii + 338. $4.50.

Schachtel, Hyman J. *The real enjoyment of living.* New York: Dutton, 1954. Pp. 192. $2.75.

Smith, Charles R. *The mind and the universe.* New York: William-Frederick Press, 1954. Pp. 173. $3.50.

Sonnemann, Ulrich. *Existence and therapy.* New York: Grune & Stratton, 1954. Pp. xi + 372. $5.00.

BUREAU OF EDNL. & PSYL.

Authors of Books and Tests Reviewed

Abrahams, J., 151
Ainsworth, Mary D., 230
Anastasi, Anne, 472
Andrew, G., 475
Arnold, Magda B., 384
Auld, F., Jr., 74

Bach, G. R., 229
Beck, S. J., 472
Bellsmith, Virginia, 386
Bennett, E. M., 387
Berg, I. A., 308
Betts, G. L., 77
Blum, Lucille Hollander, 309
Brown, J. S., 74
Brown, W. F., 153
Burgemeister, Bessie B., 309
Burington, R. S., 384
Burt, C., 473

Cantor, N., 151
Carmichael, L., 307
Christie, R., 229
Coleman, Marie L., 152
Cronbach, L. J., 473

Daniel, R. S., 74
Doll, E. A., 74
Dollard, J., 74
Dvorine, I., 154

Eisenson, J., 309
Eissler, Ruth S., 229
English, H. B., 229
English O. S., 473
ysenck, H. J., 75, 229

Festinger, L., 75
Finch; F. H., 77
Finch, S. M., 473
Fink, H. K., 384
lanagan, J. C., 231
ank, J. D., 76
ank, L. K., 229
ud, S., 151,384

son, J. A., 384
bhard, P. H., 75
rken, C. d'A., 230
l, M., 385
nsburg, S. W., 151
inzberg, E., 151
ladwin, T., 307
ordon, L. V., 154
reenacre, Phyllis, 307

hn, M. E., 230
rlow, H. F., 74
rms, E., 308

Hartley, E. L., 229
Hartwell, S. W., 475
Hathaway, S. R., 75
Herma, J. L., 151
Hertzman, M., 231
Hodgson, K., 473
Holt, R. R., 230
Holtzman, W. H., 153
Hovland, C. I., 152
Humphreys, A., 385
Hurlock, Elizabeth B., 473
Hutt, M. L., 475

Ingham, H. V., 230
Ingram, Christine P., 152

Jahoda, Marie, 229
Janis, I. L., 152

Katz, D., 75
Kelley, H. H., 152
King, J. E., 77
Kinsey, A. C., 75
Klopfer, B., 230
Klopfer, W. G., 230
Kuhlmann, F., 77

Lazarsfeld, P. F., 473
Lehman, H. C., 76
Leuba, C., 229
Lewis, H. B., 231
Libo, L. M., 230
Loewenstein, R. M., 152
Lorand, S., 474
Lorge, I., 309
Louttit, C. M., 74
Love, Leonore R., 230

Machover, K., 231
Martin, C. E., 75
May, D. C., Jr., 384
Meissner, P. B., 231
Monachesi, E. D., 75
Mowrer, O. H., 74

Newcomb, T. M., 74
Newman, R., 385
Nichtenhauser, A., 152
Notcutt, B., 153
Nowlis, V., 74

Paterson, D. G., 230
Peck, H. B., 386
Peck, L., 76
Pennington, L. A., 308
Pepinsky, H. B., 308
Pepinsky, Pauline Nichols, 308
Piaget, J., 76
Pickford, R. W., 474
Pomeroy, W. B., 75

Postman, L. J., 74
Powdermaker, Florence B., 76

Redlich, F. C., 385
Ruhe, D. S., 152

Sarason, S. B., 307, 474
Sargent, Helen D., 76
Schilder, P., 153
Schlosberg, H., 386
Sommers, Margaret, 385
Stolurow, L. M., 153
Stolz, Lois Meek, 474
Stott, D. H., 308
Sullivan, H. S., 475
Swartz, H., 475

Traxler, A. E., 385
Tyler, Leona E., 231

Varon, Edith, 151
Vedder, C. B., 308
Vernon, P. E., 153
Viteles, M. S., 77

Walton, R. E., 475
Wapner, S., 231
Warters, Jane, 386
Wesman, A. G., 309
Westwood, G., 475
White, Alice Marsden, 74
Wingert, Judith W., 77
Witkin, H. A., 231
Wolberg, L. R., 386
Wolffheim, Nelly, 308
Wolstein, B., 386
Woodworth, R. S., 386
Wulfeck, J. W., 387

Fl
Fr
Fr.
Fre

Gas
Ge
Ge.
Gil
Gi

Ia
Ha
Ha